LUCY MCIVOR, BSc (Hons), MSc, MPhil

THE ASSESSMENT AND TREATMENT OF 'COLD' EXECUTIVE FUNCTIONING IMPAIRMENTS

Section A: A Systematic Review of Interventions for Cold Executive Functioning Impairments amongst Adults with Acquired Brain Injury
Word count*: 7,961 (208)*

Section B: Increasing the Clinical Utility of the SPANS-X Conceptual Flexibility Index: A Pilot study with Healthy Adult Participants
Word count: *7,925 (390)*

Overall word count: *15,886 (598)*

A thesis submitted in partial fulfilment of the requirements of
Canterbury Christ Church University for the degree of
Doctor of Clinical Psychology

APRIL 2024

SALOMONS INSTITUTE
CANTERBURY CHRIST CHURCH UNIVERSITY

Acknowledgements

I would like to thank my participants for giving up their time to take part in in this research. I would like to thank my supervisor Dr Gerald Burgess for his unwavering support and enthusiasm. Thank you for always making time for this work, for listening to and valuing my opinions and for encouraging me to think widely and deeply about neuropsychological ideas; I have learnt so much throughout this research process. I would finally like to thank my family and friends for their continued support, encouragement, and patience.

Summary

Section A: Section A is a systematic review that investigated interventions aiming to improve cold executive functioning impairments in working age adults with an acquired brain injury. It describes interventions delivered, intervention outcomes, outcome measures used, levels of evidence, study quality and participant characteristics. The results from the review suggest that the most efficacious interventions for cold executive functioning impairments are targeted at a specific sub-construct of cold executive functioning and are driven by a neuropsychological theory that accounts for the role of attention in cold executive functioning.

Section B: Section B is a pilot study that aimed to improve the clinical utility of the conceptual flexibility index (CFI) of the Short Parallel Assessments of Neuropsychological Status-X (SPANS-X) via the addition of theoretically-driven items and subtests. The CFI of the SPANS-X was designed to assess for impairments in conceptual flexibility (a sub-construct of cold executive functioning) amongst adults with an acquired brain injury. Fifty healthy participants completed an expanded version of the CFI on Psytoolkit. Results showed that certain items and subtests showed good validity and reliability with the original version of the CFI, meriting their inclusion in further test validation studies.

Contents

Section A

Tables

Section A

Section B

Figures

Section A

Section B

Section C

LUCY MCIVOR, BSc (Hons), MSc, MPhil

THE ASSESSMENT AND TREATMENT OF 'COLD' EXECUTIVE FUNCTIONING IMPAIRMENTS

Section A: A Systematic Review of Interventions for Cold Executive Functioning Impairments amongst Adults with Acquired Brain Injury

Word count: *7,961 (208)*

A thesis submitted in partial fulfilment of the requirements of
Canterbury Christ Church University for the degree of
Doctor of Clinical Psychology

APRIL 2024

SALOMONS INSTITUTE
CANTERBURY CHRIST CHURCH UNIVERSITY

Selected journal: Neuropsychological Rehabilitation

Abstract

Objective: This systematic review investigated interventions that aimed to improve cold executive functioning in working age adults with an acquired brain injury (ABI).

Methods: The following databases were searched: PsycINFO, Web of Science, CENTRAL, Medline, CINAHL and EMBASE. The search was limited to peer-reviewed research studies published between January 2000-October 2023. Articles were classified according to the Oxford Centre for Evidence-Based Medicine levels of evidence criteria. Joanna Briggs Institute (JBI) appraisal checklists were used to measure study quality.

Results: This systematic review reported on 12 studies. Interventions delivered, intervention outcomes, cold executive function outcome measures used, levels of evidence, study quality and participant characteristics are described.

Conclusion: Few methodologically high-quality studies exist in this research area, so conclusions drawn regarding intervention efficacy are limited. Results showed that the most efficacious interventions were targeted at a specific sub-construct of cold executive functioning and driven by a neuropsychological theory which accounted for the role of the cognitive process of attention in cold executive functioning. In total, 25 different outcome measures were used to measure change in cold executive functioning. Outcome measures that focus on a specific component of the cold executive functioning construct may benefit both future research and clinical practice.

Keywords: Brain injuries, executive functioning, attention

Introduction

Defining Acquired Brain Injury (ABI)

Acquired brain injury (ABI) is defined as any form of non-degenerative brain injury sustained after birth (Beecham et al., 2008). In 2019, there were 1.4 million individuals in the United Kingdom living with an ABI (House of Commons, 2019). The term encapsulates both traumatic brain injury (TBI) and non-traumatic brain injury. Causes of TBI include motor accidents, sports-related injuries, falls and being the victim of violence. Causes of non-traumatic ABI include stroke, tumour, infection, anoxia and surgical complications (Bruns & Hauser, 2003; Colantonio et al., 2011).

Cold Executive Functioning Impairment Following ABI

Executive functions are defined as higher-order cognitive abilities that enable individuals to perform goal-directed tasks, cope and adapt to novel situations, and navigate complex scenarios in the course of everyday functioning (Diamond, 2013; Suchy, 2009).

One well accepted theoretical framework for investigating executive functions is to distinguish between cold and hot components (Ward, 2020). Within this framework, cold executive functions are designated as decontextualised processes recruited under analytic and non-emotional testing conditions, whilst hot executive functions are designated as affective and related to the processing of information relevant to reward, emotion, and motivation (Peterson & Welsh, 2014; Salehinejad et al., 2021). This study focuses on cold executive functions specifically.

Impairments in cold executive functioning are one of the most frustrating potential consequences of ABI. Such impairments may present as difficulties with starting or completing tasks, multi-tasking, making decisions, thinking through problems, forming solutions, planning and organisation (Colautti et al., 2022; De Luca & Leventer, 2008). To cope with these

difficulties, individuals may adhere to rigid routines, showing a lack of flexibility in response to everyday situations (Perna et al., 2012). Impairments in cold executive functioning can therefore cause profound challenges, impacting on an individual's functional and vocational abilities (Chan et al., 2008; McDonald et al., 2002).

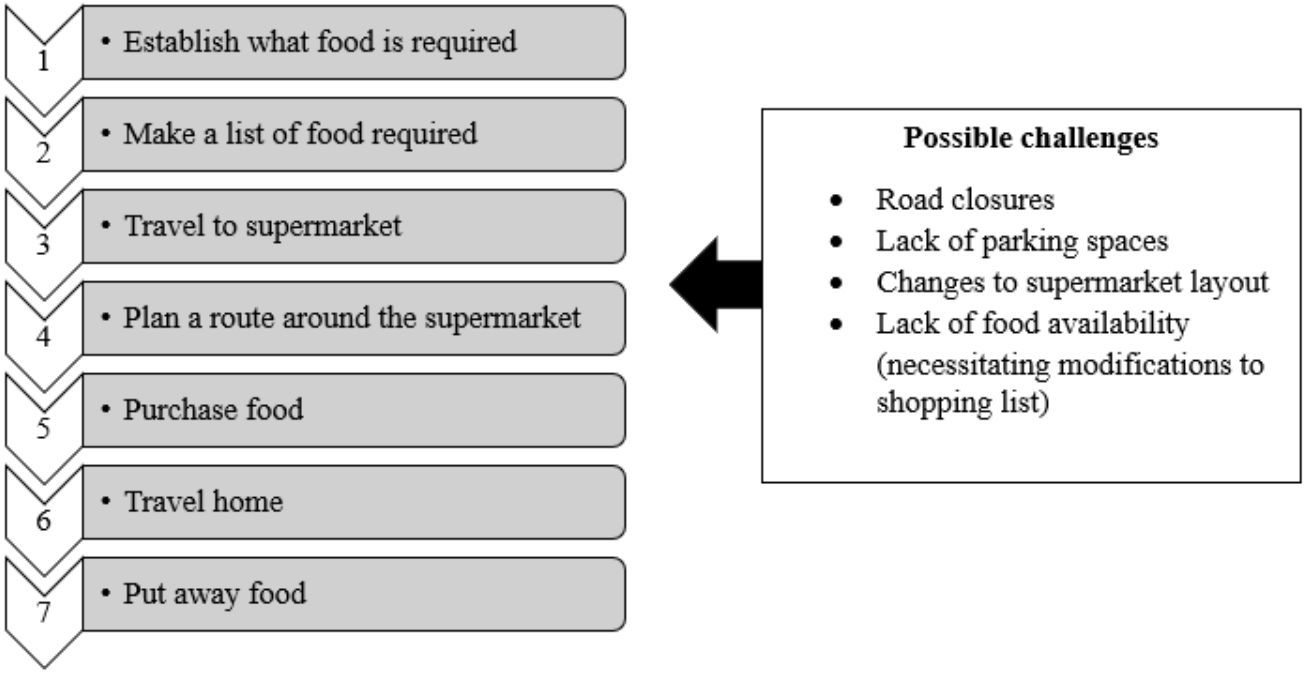
Impact on Functional Abilities

Functional abilities include activities of daily living (ADLs) that may be split into basic activities of daily living (BADLs) and instrumental activities of daily living (IADLs) (Lawton & Brody, 1969). BADLs include activities such as feeding, toileting, bathing, and dressing. IADLs include activities such as grocery shopping, meal preparation, telephone use, financial management and driving.

IADLs present challenges for individuals with impairments in cold executive functioning for several reasons. First, unlike more basic activities, IADLs require a greater level of neuropsychological organisation to carry out, as they consist of sequential 'steps', each of which must be completed to achieve the end goal of the task (Wesson et al., 2016) (Figure 1). The sequential completion of these 'steps' is likely to be challenging for individuals with impairments in cold executive functioning, due to difficulties with planning and organisation. Second, unlike BADLs, IADLs are less 'reliable', less protocolised and more prone to change (Figure 1). Any changes necessitate a 'flexible response' which is likely to be difficult for those with impairments in cold executive functioning, due to rigid reliance on predetermined routines as a coping strategy. In some cases, individuals may be unable to complete IADLS without significant additional assistance, particularly when an individual's required level of social support is not available (Perna et al, 2012).

Figure 1

Sequential Steps Required to Complete the Activity of Grocery Shopping



Impact on Vocational Abilities

Returning to work poses a particular challenge for those who have sustained an ABI, as most forms of employment require grossly intact cold executive functioning (Shames et al., 2007). Though workplaces are legally required to implement adjustments for those returning to work after ABI, the burden of injury is often underestimated and misunderstood by employers (Stergiou-Kita et al., 2017). In addition, many workplaces require employees to work within time constraints to ensure that deadlines are met, meaning the compensatory strategy of taking extra time to complete tasks is excluded from the coping repertoire of ABI sufferers. Research conducted on employment outcomes amongst those with ABI indicated that only 37% of individuals were able to return to the same form of work undertaken before injury. Sixteen percent of individuals returned to work at a lower level, 28% retired within a period of two years after restarting work, and 19% reported persistent difficulties in maintaining and keeping a job. These findings suggest that under half of individuals who have sustained an ABI are able

to return to employment, even if the work being undertaken represents a ‘lower level’ as compared to before the injury (Pössl et al., 2001). An inability to engage with employment is likely to have detrimental emotional, social, and financial consequences for ABI sufferers and their families (Coetzer et al., 2011).

Variation in Cold Executive Functioning Impairment following ABI

The nature and degree of cold executive functioning impairment experienced as a result of ABI varies widely between individuals. Impairments may be specific or widespread, and temporary or permanent (Ownsworth & Clare, 2006). Impairments also vary in the extent to which they change over time, in that whilst some individuals will fully recover their cold executive functioning, others will be left with life-long difficulties (Barman et al., 2016). The nature and degree of impairment is influenced by individual factors such as age and premorbid functioning (Verdugo et al., 2019), injury related factors such as the location, extent and severity of ABI (Williams et al., 2015) and external factors such as the quality of care and rehabilitation received in the aftermath of sustaining an ABI (Downing et al., 2021).

Interventions for Cold Executive Functioning Impairments following ABI

The frustrating consequences of cold executive functioning impairment following ABI highlight the need for efficacious interventions. A range of interventions aimed at improving cold executive functions amongst those who have sustained an ABI currently exist. The nature of these interventions tends to differ depending on their intended timing of delivery after ABI. Early-stage interventions delivered in the acute phase after ABI often focus on restoring as much cold executive functioning as possible. This is because the majority of spontaneous recovery occurs during this phase, thus natural repair processes aid external rehabilitation. In contrast, later stage interventions often aim to optimise a person’s residual level of executive functioning through more compensatory style training. This is because recovery gains during this time period are more marginal (Barman et al., 2016; Holleman et al., 2018).

Current interventions to improve cold executive functioning following ABI include targeted cognitive training and rehabilitation programmes (e.g. Krasny-Pacini et al., 2014), as well as more complementary therapeutic approaches such as music training (e.g. Sihvonen et al., 2017). Following the COVID-19 pandemic, an increasing number of interventions incorporate some form of computer-based learning or virtual reality. To the authors knowledge, no systematic review of interventions for cold executive functions amongst adults with ABI has been undertaken. The efficacy of these interventions for this population is therefore not known.

Executive Functioning: Theoretical Considerations

It is important to note for the purposes of this review that the complexity of 'executive functioning' as a construct has rendered a universally accepted definition impossible, and various theoretical debates continue within the scientific community. Whilst a full exploration of these debates is beyond the scope of the review, it is important to highlight certain aspects for the purpose of clarity.

First, a convergence of evidence suggests that the high-order cognitive abilities that constitute executive functioning (e.g. problem-solving) recruit a range of more basic cognitive systems (e.g. attention and memory) as part of their mechanism of action (Kennedy et al., 2008). As an example, in order to problem-solve, one must have the memory required to hold the problem in mind and the attention required to attend to the problem. What is not known is the extent to which these more basic cognitive systems interact with each other (Suchy, 2009).

Second, whilst a range of neuropsychological tests and outcome measures purport to measure executive functioning (See Burgess [2010] for more details), it seems to some extent that executive functioning has been operationalized by the tasks used to measure it. This circular reasoning has borne neuropsychology's own chicken-or-egg question as to what came first, the construct of executive functioning or its invented tasks. This debate has drawn limited

conclusions and the existing body of literature on executive functioning shows high levels of heterogeneity both in how executive functioning is defined and how it is measured (Cicerone, 2002).

Rationale

Impairments in cold executive functioning as a result of ABI are a source of clinical relevance and concern. To the author's knowledge, no systematic review of interventions for cold executive functions amongst adults with ABI has been undertaken. Those which do exist have focused on TBI specifically and have explored executive functioning as a single domain (Kennedy et al., 2008). A systematized evaluation of interventions for cold executive functions amongst adults with ABI may lead to useful clinical recommendations in terms of the most efficacious interventions.

Aims

This review aims to answer the following research questions:

1a) What is the efficacy of interventions aimed at improving performance on cold executive functioning outcome measures amongst adults with ABI?

1b) Do the more efficacious interventions share any characteristics?

Method

Selection of Studies

The systematic review was registered on Prospero prior to data extraction. The systematic review was conducted in line with the Prisma Checklist (Page et al., 2021, Appendix A). The following databases were searched: PsycINFO, Web of Science, CENTRAL, Medline, CINAHL, and EMBASE. The search was limited to peer-reviewed research studies published between 1st January 2000-26th October 2023 in the English language via database filters. The rationale for this time frame was that the year 2000 broadly coincides with the 'digital age' (Hilbert, 2020) and a paradigm shift within neuropsychology, including the digitalisation of neuropsychological tests and interventions.

When searching the PsycINFO database, Medical Subject Headings (MeSH) were mapped onto the original search terms and grouped as population, impairment and intervention type. Appendix B outlines the keywords used for the original and MeSH systematic search.

In accordance with previous research (Chavez-Arana et al., 2018; De Luca and Leventer 2008), cold executive functions were defined as 'mental flexibility', 'cognitive flexibility', 'conceptual flexibility', 'metacognition', 'decision-making', 'planning and problem-solving' (Appendix B, list A). Hot executive functions were defined as 'behaviour-regulation', 'emotion-regulation', 'affective decision-making', 'social skills' and 'theory of mind'. Hot executive functions were excluded from this review.

The rationale for excluding hot executive functions from this review is that the construct of 'executive functioning' is exceedingly broad, and there is evidence to suggest that hot and cold executive functions are neuro-geographically and theoretically opposed (Salehinejad et al., 2021). It logically follows that the most efficacious interventions for cold executive functioning impairments may be different to those for hot executive functioning impairments. Therefore,

the inclusion of cold executive functions only allowed for focus on a subset of conceptually similar executive functions to ensure that the context of the review was not too generalized to be meaningful.

Studies that investigated the cognitive processes from list B (Appendix B) were not included in the review if the cognitive processes were investigated in isolation (e.g. a study exploring attention with outcome measures related to attention only). This decision was informed by previous research (Kennedy et al., 2008) that indicated that the list B search terms are more basic cognitive processes that exist in their own right. They therefore do not constitute cold executive functions, which operate via a combination of these more basic processes. It was, however, important to include the search terms from list B within the systematic search because certain study titles included search terms from list B but not list A but were nevertheless eligible for inclusion on account of having an outcome measure related to cold executive functioning. For example, the study by Yoshida et al. (2018) entitled 'Flow experience enhances the effectiveness of attentional training: a pilot randomized controlled trial of patients with attention deficits after traumatic brain injury', was eligible for inclusion on account of stating that one outcome measure was intended to measure cold executive functioning, the Wisconsin Card Sort Test ([WCST], Grant & Berg, 1948), despite this not being alluded to in the title.

List A (Appendix B) also contained the search term 'executive function'. This search term captured studies with outcome measures related to both hot and cold executive functions, and also studies with outcome measures related to hot executive functions only. Therefore, when the term 'executive function' appeared in the title of studies, these studies were read in full, and only outcome measures related to cold executive functions were reported and synthesised as part of this review. For example, the included study by Fong and Howie (2009) incorporated seven outcome measures: The Key Search Test and the Modified Six Elements Test from The Behavioural Assessment of the Dysexecutive Syndrome ([BADS] Wilson et al., 1996) and the

Metacomponential Interview [MI], Clements & Nastasi, 1990) to measure problem-solving, the Social-Problem-Solving Video Measure ([SPSVM], Kendall et al., 1997) and the Means-End-Problem-Solving-Measure ([MEPSM], Spivack et al., 1976) to measure social problem-solving, Raven's Progressive Matrices ([RPM], Raven et al., 2000) to measure inductive reasoning, and the Rivermead Behavioural Memory test ([RMBT], Ng et al., 1998) to measure everyday memory ability. The study was included on account of having three outcome measures related to cold executive functioning: The Key Search Test and the Modified Six Elements Test from the BADS, and the MI. The results of these outcome measures are therefore synthesised (Table 2, Table 3). The outcome measures related to social problem-solving are not synthesised, as social problem-solving is a hot executive function. The outcome measures related to inductive reasoning and everyday memory ability are not synthesised, as they are not cold executive functions but represent more basic cognitive processes.

Requirements for Inclusion

After duplicates were excluded, abstracts were reviewed by the author to determine eligibility for inclusion (Table 1).

Table 1

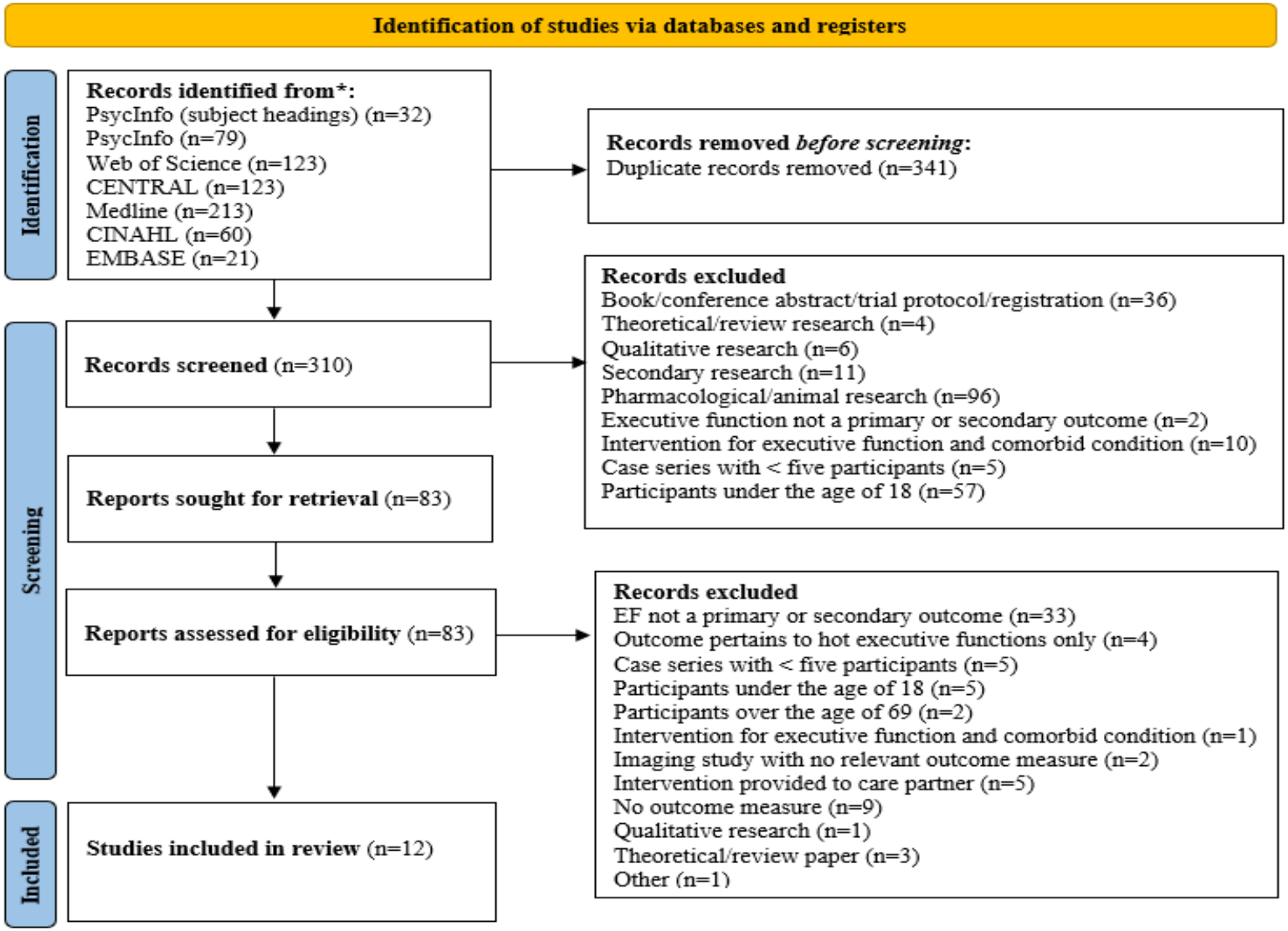
Inclusion and Exclusion Criteria

Inclusion criteria	Exclusion criteria
• Participants of working age (aged 18-69, [Department of Health, 2017]) • Any non-pharmacological intervention study aiming to improve cold executive functioning in participants with ABI • Cold executive functions measured as either a primary or secondary outcome, either by rating checklist, questionnaire, or performance-based task • At least two measurement points, before and after intervention	• Theoretical/review/qualitative studies • Neuroimaging studies that did not also measure cold executive functioning as a primary or secondary outcome measure • Pharmacological/animal-based studies • Case studies with under five participants • Studies that explicitly aimed to investigate the effect of an intervention on both executive functioning and a comorbid psychological difficulty (e.g. investigating the efficacy of a rehabilitation strategy on both executive functioning and binge drinking behaviour concurrently) • Studies where interventions were provided to the care partner or families of individuals with ABI but not the individuals themselves

The adapted Prisma flow diagram (Page et al., 2021) in Figure 2 illustrates the selection process and the results of the systematic search. A total of 310 studies were screened from which 227 were excluded based on abstract alone. The full text of 83 studies were read to determine eligibility. The final review reported on 12 studies.

Figure 2

Prisma Flow Diagram Illustrating the Selection Process and Results of the Systematic Search



Classification of the Studies and Quality Grading

All articles were classified according to the Oxford Centre for Evidence-Based Medicine (2009) levels of evidence criteria (Appendix C). The quality of each study was assessed using the Joanna Briggs Institute (JBI) appraisal tools for randomised controlled trials (RCTs) (Barker et al., 2023), quasi-experimental studies (Tufanaru et al., 2017) and case-series (Munn et al., 2020) as appropriate. These checklists were used as they enabled a uniform evaluation approach to be taken across the different study designs included in this review. The JBI checklists assess bias across two domains, internal validity and statistical conclusion validity. The JBI checklist for RCTs contains ten questions related to internal validity and three questions related to statistical conclusion validity (Appendix D). This checklist was used to evaluate 10 studies. The JBI checklist for quasi-experimental designs contains nine questions (Appendix E). This checklist was used to evaluate one study. The JBI checklist for case-series contains 10 questions (Appendix F). This checklist was used to evaluate one study.

As not all checklists contained a question pertaining to the validity and reliability of the outcome measures used, questions were added to each checklist by the author to provide this additional information when assessing risk of study bias. These additions are summarised in Appendix G. It should be noted that whilst the JBI checklist for RCTs includes a question pertaining to reliability (were outcomes measured in a reliable way?), this question referred specifically to rater reliability, specifically the number and training of outcome raters. An additional reliability question was therefore added to address the reliability of the outcome measures specifically (were outcomes measured in a reliable way? [test reliability]). The additions increased the number of questions on the JBI checklists for RCTs to 15, the number on the JBI checklist for quasi-experimental studies to 10, and the number of questions on the JBI checklist for case series to 12.

Data Synthesis

A score for overall adherence to the JBI checklist criteria (yes/no/unclear for each question) was calculated for each study. The answer 'yes' was allocated a score of one, and the answers 'unclear' and 'no' were allocated a score of zero. Reports of effect sizes were generally lacking in the shortlisted articles, and the outcome measures used to measure cold executive functioning performance were heterogeneous. Therefore, the approach of a previous systematic review (Chavez-Arana et al., 2018) was followed, and improvement after an intervention was defined as $p < 0.05$ as reported by the authors on at least one relevant (related to cold executive functioning) outcome measure. Due to the data heterogeneity and the lack of reporting regarding effect sizes, conducting a meta-analysis was not possible. However, the % of studies targeting cold executive functioning that found improvement on at least one relevant outcome measure are reported in Appendix H.

Results

Nature of Interventions for Cold Executive Functioning

The cold executive functions investigated by the studies reviewed were executive functioning (four studies), problem-solving (five studies), mental flexibility (two studies), and a composite measure of attention and executive function (two studies). The interventions delivered across the studies varied widely. Ten studies delivered specific cognitive training programmes, one study delivered 'flow' training to induce a psychological 'flow state' (Yoshida et al., 2018), and one study delivered a music intervention with one session focused on improving executive functioning impairments specifically (Thaut et al., 2009). The delivery mode of interventions varied between studies. Six studies reported exclusively face-to-face interventions, five studies reported interventions that were a combination of face-to-face and online, and one study reported an exclusively online intervention. The duration and frequency of sessions also varied between studies. Three studies reported on the duration and frequency of sessions, four studies reported either duration or frequency of sessions, and four studies reported no information. A further study (Fong & Howie, 2009) reported the duration and frequency of sessions for the experimental group but not the control group. The minimum reported duration of sessions was four weeks, and the maximum reported duration of sessions was 24 weeks. The minimum reported frequency of sessions was once every second week, and the maximum reported frequency of sessions was three to four times a week. The number of sessions provided to participants and the time length of these sessions also varied between studies. Eight studies reported the number of sessions provided to participants and the time length of these sessions. Two studies reported the number of sessions only, and one study reported no information. A further study (Fong & Howie, 2009) reported the number of sessions and the time length of sessions for the experimental group but not the control group. The minimum reported number of sessions was four sessions, and the maximum reported number of sessions was 40 sessions.

The minimum reported time length of sessions was 20-25 minutes whilst the maximum reported time length of sessions was 120 minutes. All studies listed the job titles of the intervention providers. A range of different job titles were listed including examiners, therapists, occupational therapists, graduate/certified speech and language pathologists, psychologists, neuropsychologists, board certified music therapists and vocational therapists (Table 2).

Outcome of Interventions for Cold Executive Functioning

There were three possible outcomes for the studies reviewed; cold executive functioning improved in response to a specific intervention, a generalised improvement in cold executive functioning over time occurred regardless of intervention, or no change in cold executive functioning occurred (Table 2). Table 2 is supplemented by Table 3, which provides in-depth information regarding the outcome measures used to assess cold executive functioning. Two studies reported no change in cold executive functioning. Two studies reported a generalised improvement in cold executive functioning over time regardless of intervention, suggesting that, for studies with more than one intervention, the nature of the intervention had little influence on cold executive functioning improvement. It should be noted that one of these studies (Marshall et al., 2004) followed a case-series design so only one intervention was delivered. Eight studies reported improvement in cold executive functioning in response to a specific intervention. For seven out of eight of these studies, improvement was observed in the intervention as opposed to the comparison group (see Table 3 for further details). For one study only (Fong & Howie, 2009) the comparison group performed better than the experimental group, on the Key Search Test of the BADS from post-intervention to follow-up. Appendix H details the % of studies targeting cold executive functioning that found improvement on at least one relevant outcome measure.

Table 2

Interventions that Aim to Improve Cold Executive Functioning

First author (year)	Level of evidence	JBI score	Relevant target of the intervention	Intervention delivered	Delivery mode	Duration of sessions (frequency)	Number of sessions (time)	Job title of intervention providers	Attrition
Soong (2005)	2b	3/15	Problem-solving (*)	1. Online programme 2.Computer assisted programme 3.Therapist-administered programme	1. F2F 2.Online 3. F2F	-	20 sessions -	Therapists	-
Yoshida (2018)	2b	11/14	Executive function (0)	1.Flow video game 2.Control video game	1.Online 2.Online	4 weeks -	40 sessions -	NA (computer games)	5.00%
Jacoby (2013)	2b	6/15	Executive function (0)	1.Cognitive retraining 2.Cognitive retraining + Virtual Mall (VMall)	1.F2F 2. F2F + online	- (3-4 times a week)	10 sessions (45 minutes)	Occupational therapists	7.69%
Fong (2009)	2b	7/15	Problem-solving (+)	1.Functional skills training 2.Functional skills training + Metacomponential skills training	1.F2F + online 2.F2F + online	Intervention group: 15 weeks (twice a week)	Intervention group: 22 sessions (65 minutes)	Occupational therapists	33.33%
Marshall (2004)	4	8/12	Problem-solving (*)	1.Interactive StrategyModelling Training (ISMT)	1.F2F	-	-	Examiners	0.00%

Sohlberg (2000)	2b	5/15	Executive function (+)	1.Attention Process Training (APT) 2.Therapeutic support condition	1.F2F + online 2.F2F	10 weeks (once a week)	Intervention group: 10 sessions (180 minutes) Control group: 10 sessions (60 minutes)	Clinicians: Certified speech/language pathologists or graduate students in speech/ language pathology	14.29%
Rath (2003)	2b	5/15	Problem-solving (+)	1.Conventional group neuropsychological rehabilitation 2.Innovative treatment focused on problem-solving deficits	1.F2F 2.F2F	24 weeks (once a week)	24 sessions (120-180 minutes)	Psychologists	23.33%
Thaut (2009)	2c	5/10	Mental Flexibility (+)	1.Musical intervention 2.Rest	1.F2F 2.F2F	-	4 sessions (30 minutes)	Board certified music therapist	-
Tornås (2016)	2b	5/15	Executive function (+)	1.Goal Management training (GMT) 2.Brain Health Workshop (BHW) (psycho-educative active control condition)	1.F2F 2.F2F	8 weeks (2 sessions per day 1 day every second week)	8 sessions (120 minutes)	Neuropsychologist (also primary investigator of study)	4.29%

Man (2013)	2b	7/15	Problem-solving (+)	1.Artificial Intelligent Virtual Reality-based Vocational Training System (AIVTS) 2.Psycho-educational vocational training programme	1.Online 2.F2F	-	12 sessions (20-25 minutes)	1. NA (virtual reality) 2. Vocational trainer	20.00%
Novakovic-Agopian (2018)	2b	10/15	Attention and executive function (+) Mental flexibility (+)	1.Goal-Oriented Attentional Self-Regulation (GOALS) 2.Brain-Health Education (BHE)	1.F2F 2.F2F	5 weeks -	10 sessions group training (120 minutes) + 3 individual sessions (60 minutes) + 20h at home practice	Therapists	5.71%
Chen (2011)	2b	6/15	Attention and executive function (+)	1. Goal-Oriented Attentional Self-Regulation (GOALS) 2. Education control condition	1.F2F 2.F2F	5 weeks -	10 sessions group training (120 minutes) + 3 individual sessions (60 minutes) + 20h at home practice	Occupational therapists Neuropsychologists	0%

Note. - = unclear/not provided (+) = significant intervention improvement, (*) = significant time improvement, (0) = no significant improvement.

Outcome Measures used to Measure Cold Executive Functioning Improvements

A total of 25 different outcome measures were used to measure improvements in cold executive functioning by the authors (Table 3). The WCST was used in three different studies. The following outcome measures were used in two different studies; the Trail Making Test of the Halstead-Retain Test Battery ([HRTB], Reitan, 1958), the Verbal Fluency Test from the Delis-Kaplan Executive Functioning System ([D-KEFS],Delis et al., 2001), Auditory Consonant Trigrams (Stuss et al., 1998), and the Letter Number Sequencing subtest from the Wechsler Adult Intelligence Scale III ([WAIS-III, Wechsler, 1997). Two studies (Chen et al., 2011; Novakovic-Agopian et al., 2018) used composite scores complied of outcome measures from various cognitive domains to measure cold executive functioning outcomes.

Table 3

Detailed Test Characteristics and Outcome Measures of Included Studies

First author (year)	Relevant target of the intervention	Relevant test name	Aspect of test used	Significance
Soong (2005)	Problem-solving	1. Category Test of the Halstead-Reitan Test Battery (HRTB) (Reitan & Wolfson, 1958) 2. Lawton Instrumental Activities of Daily Living (IADL) (Lawton & Brody, 1969) 3. Self-efficacy checklist (Soong et al., 2005)		1.+ (time) 2.+(time) 3.0
Yoshida (2018)	Executive function	1. Wisconsin Card Sort Test (WCST) (Grant & Berg, 1948)	1a. Categories achieved 1b. Perseverative errors in Milner (PEM) 1c. Perseverative errors in Nelson	1a. 0 1b. 0 1c. 0
Jacoby (2013)	Executive function	1. Multiple Errands Test-Simplified Version (MET-SV) (Alderman et al., 2003) 2. Executive Function Performance Test (EFPT) (Baum et al., 2007)		1.0 2.0
Fong (2009)	Problem-solving	1. Behavioural Assessment of The Dysexecutive syndrome (BADS): Key Search Test (Wilson et al., 1996) 2. Behavioural Assessment of The Dysexecutive syndrome (BADS): Modified Six Elements Test (Wilson et al., 1996) 3. Metacomponential interview (MI) (Clements & Nastasi, 1990)	3a. Nature (Cor) 3b. Nature (Meta) 3c. Planning (Cor) 3d. Planning (Meta) 3e. Representation (Cor) 3f. Representation (Meta) 3g. Monitoring (Cor)	1. 0 2. + (Comparison of change scores time 2-time 3) superior = functional skills training (comparison group) 3a. 0 3b. 0 3c. 0 3d. 0

				3h. Monitoring (Meta) 3i. Total (Cor) 3j. Total (Meta)	3e. + (Comparison of change scores time 1-time 2) superior = functional skills training + Metacomponential training (experimental group) 3f. 0 3g. 0 3h. 0 3i. + (Comparison of change scores time 1-time 2) superior = functional skills training + Metacomponential training (experimental group) 3j. 0
Marshall (2004)	Problem-solving	1. Rapid Assessment of Problem-Solving Test (RAPS) (Marshall et al., 2003)		1a. Number of questions 1b. % of Constraint seeking questions 1c. Question-asking efficiency scores	1a. + (Time) 1b. + (Time) 1c. + (Time)
Sohlberg (2000)	Executive function	**1. Executive function (composite score)** 1. Trail Making Test of the Halstead-Reitan Test Battery (HRTB) (Reitan, 1958) 2. Stroop Task (Stroop, 1935)		 1a. Part A 1b. Part B 2a. Error interference	**1. + (intervention received, intervention received x vigilance) superior = APT (experimental group) x high vigilance**
Rath (2003)	Problem-solving	1. The Wisconsin Card Sort Test (WCST) (Grant & Berg, 1948)		1a. Perseverative response score	1a. + (Time for group one) Superior = innovative group (experimental) focused on problem-solving deficits
Thaut (2009)	Mental flexibility	1. Trail Making Test of the Halstead-Reitan Test Battery (HRTB) (Reitan, 1958)		1a. Part B	1a. + (Time for group two) superior = music group (experimental)

Tornås (2016)	Executive function	1. Conner's Continuous Performance II (Conners, 2000) 2. Delis-Kaplan Executive Function System (DKEFS) Color-Word Interference Test (CWI) (Delis et al., 2001) 3. Delis-Kaplan Executive Function System (DKEFS) Verbal Fluency Test (Delis et al., 2001) 4. Delis-Kaplan Executive Function System (DKEFS) Tower Test (Delis et al., 2001) 5. The Hotel Task (Manly et al., 2002) 6. The UCSD Performance-Based Skills Assessment (Patterson et al., 2001)	1a. Omission errors 1b. Commission errors 1c. Reaction time 2a. Color-Word Interference Test (CWI)-inhibitory control 2b. Color-Word Interference Test (CWI)-switching 3. Verbal fluency test (VFT) sum of category and repetition errors 4a. The Tower Test: number of rule violations 4b. The Tower Test: total Time 4c. The Tower Test: total achievement score 5a. No. of tasks attempted 5b. Deviation from optimal time (s)	1a.+(Time) 1b.+(Time) 1c.0 2a. + (Intervention received x time) superior = GMT (experimental group) 2b. 0 3. + (Intervention received x time) superior = GMT (experimental group) 4a. + (Time) 4b. + (Time) 4c. + (Time) 5a. + (Time) 5b. + (Time) 6.+ (Time)
Man (2013)	Problem-solving	1. Tower of London (TOL) (Shallice, 1982) 2. Wisconsin Card Sorting Test (WCST)-Computer version 4 (Grant & Berg, 1948) 3. The Vocational Cognitive Rating Scale (VCRS) (Greig et al., 2004)	2a. % errors 2b. % perseverative errors 2c. Conceptual level response	1. 0 2a. + (Intervention received x time) superior = Artificial intelligent virtual reality-based

				vocational training system (experimental) 2b. 0 2c. + (intervention received x time) superior = Artificial intelligent virtual reality-based vocational training system (experimental) 3.0
Novakovic-Agopian (2018)	Attention and executive function Mental flexibility	**1. Attention and executive function (composite score)** 1. Wechsler Adult Intelligence Scale III (Wechsler, 1997) 2. Auditory Consonant Trigrams (Stuss et al., 1998) 3. Delis-Kaplan Executive Function System (DKEFS) Stroop inhibition (Delis et al., 2001) 4. Delis-Kaplan Executive Function System (DKEFS) Design Fluency Test (Delis et al., 2001) 5. Delis-Kaplan Executive Function System (DKEFS) Verbal Fluency Test (Delis et al., 2001) 6. Delis-Kaplan Executive Function System (DKEFS) Trail Making Test (Delis et al., 2001) 7. Digit vigilance Test (Heaton et al., 2004) **2. Mental flexibility (composite score)**	1a. Letter number sequencing 3a. Stroop inhibition task time 3b. Stroop inhibition task errors 4. Design-Fluency Switching 5. Verbal Fluency-Switching 6a. Part B 7a. Time scores 7b. Error scores	**1.+ (Time, intervention received x time, Time for group one) superior = Goal Oriented Attentional Self-Regulation (GOALS) (experimental)** **2. + (Time, time for group one) superior = Goal Oriented**

		1. Delis-Kaplan Executive Function System (DKEFS) Trail Making Test (Delis et al., 2001) 2. Delis-Kaplan Executive Function System (DKEFS) Stroop inhibition (Delis et al., 2001) 3. Delis-Kaplan Executive Function System (DKEFS) Design-Fluency Switching (Delis et al., 2001) 4. Delis-Kaplan Executive Function System (DKEFS) Verbal-Fluency Switching (Delis et al., 2001)	1a. Part B 2ai. Stroop inhibition task time 2aii. Stroop inhibition task errors 3. Design-Fluency Switching 4. Verbal-Fluency Switching	**Attentional Self-Regulation (GOALS) (experimental)**
Chen (2011)	Attention and executive function	**1. Attention and executive function (composite score)** 1. Wechsler Adult Intelligence Scale III (Wechsler, 1997) 2. Auditory Consonant Trigrams (Stuss et al., 1998) 3. Digit Vigilance Test (Heaton et al., 2004) 4. Delis-Kaplan Executive Function System (DKEFS) Stroop inhibition (Delis et al., 2001)	1a. Letter number sequencing 3a. Time scores 3b. Error scores 4ai. Stroop inhibition/switching time 4aii. Stroop inhibition/switching errors 4bi. Stroop inhibition time 4bii. Stroop inhibition errors	**1.+ (Intervention received) superior = Goal Oriented Attentional Self-Regulation (GOALS) (experimental)**

Note. Measures denoted in bold are composite scores of the neuropsychological process specified. Authors calculated composite scores by combining the scores for the individual neuropsychological tests listed directly below each composite score. For example, Chen (2011) calculated an attention and executive function composite score by combining individual scores on the Wechsler Adult Intelligence Scale III, Auditory Consonant Trigrams, Digit Vigilance Test, and DKEFS Stroop inhibition.

Design of Reviewed Studies

Ten studies were classed as RCTs, one study was classed as a quasi-experimental design, and one study was classed as a case-series. Of the 10 RCT studies, two studies (Chen et al., 2011; Sohlberg, 2000) incorporated a crossover design whereby participants received both interventions in a different sequential order. Of the 10 RCT studies, seven used an active control group, whereby participants received some form of psychoeducation or support. One RCT (Soong et al., 2005) delivered three different interventions to three different groups of participants, with no clearly assigned experimental or control group. Two RCTs (Fong & Howie, 2009; Jacoby et al., 2013) took an additive approach, whereby the experimental group received the same intervention as the control group with an added interventional component. The study that followed a quasi-experimental design was the only study to have a control group that did not receive any form of intervention, as participants were assigned to 'rest' (Thaut, 2009).

Quality of Reviewed Studies

All ten RCTs were classed as level of evidence 2b (low quality RCT) due to compromised internal validity or statistical conclusion validity (Table 2). The quasi-experimental design study (Thaut et al., 2009) was classed as level of evidence 2c due to lack of randomisation, thus being classed as outcomes research. The case-series study (Marshall et al., 2004) was automatically given a level of evidence 4 on account of its design. Ten out of 12 studies reported participant rates of attrition. These ranged from 0% to 33% (Table 2).

The quality of studies reviewed varied widely as per the JBI checklists. Of the 10 studies appraised using the RCT checklists, scores ranged from 3/15-11/14 (one study was scored out of 14 not 15 as it was a computer-based study, meaning that the quality checklist question 'those delivering treatment blind to treatment?' was not applicable). The study appraised using

the quasi-experimental design checklist scored 5/10 and the study appraised using the case-series checklist scored 8/12 (Table 4).

Internal Validity of RCTs

Selection and Allocation.

All 10 studies randomly allocated participants to groups, however only four out of 10 studies reported that the allocation of participants to groups was concealed from researchers. Similarly, just four out of 10 studies evidenced that participant groups were similar at baseline prior to intervention (Table 4).

Administration of Intervention

Only two studies evidenced that participants were blinded to the intervention they received, whilst three out of nine studies evidenced that those delivering the intervention were blinded to the intervention they administered (one study involved a computer-based intervention so was not included in this count). Eight out of 10 studies evidenced that the experimental group was treated equally to the control group other than the intervention of interest, and efforts were made across the studies to match the duration of control and intervention conditions to protect validity. One study (Fong & Howie, 2009) only reported duration and frequency information for the experimental group, whilst another study (Sohlberg et al., 2000) indicated that the intervention and control group received interventions of different frequency and duration, confounding interpretation of results (Table 4).

Assessment, Detection and Measurement of Outcome.

Only five out of 10 studies reviewed indicated that outcome assessors were blinded to the interventions participants received, however all studies measured cold executive function outcomes using the same methodology for both the experimental and control groups. The reporting of interrater reliability of outcome measures was poor across studies, with just one

study (Novakovic-Agopian et al., 2018) reporting that outcome measures were assessed by more than one rater. Seven out of 10 studies evidenced validity and reliability of outcome measures, through incorporating established neuropsychological tests used to measure cold executive functioning, or by using a range of established tests across several domains to ascertain a composite measure of cold executive functioning. The remaining studies used unvalidated outcome measures, or adapted outcome measures in isolation, with poor justification as to why the outcome measures had been used to measure cold executive functioning or for what purpose (Table 4).

Participant Retention.

Six out of 10 studies either reported that all participants were retained at follow-up, or adequately described analysed differences between groups in the event of participant attrition. The remaining studies did not describe the characteristics or demographics of participants who did not complete the study (Table 4).

Statistical Conclusion Validity of RCTs

Five out of 10 studies either reported no participant attrition or incorporated 'intention to treat' (ITT) analyses to ensure results accurately reflected participant drop-out, thus avoiding bias. Only one out of 10 studies conducted a power analysis, whilst the remainder of the studies acknowledged small sample sizes as a limitation. Only one study evidenced an appropriate 'pure' RCT design, incorporating both double-blinding and ITT analyses (Tornås et al., 2016) (Table 4).

Table 4

Summary of Quality of Reviewed RCT Studies

	Soong (2005)	Yoshida (2018)	Jacoby (2013)	Fong (2009)	Sohlberg (2000)	Rath (2003)	Tornås (2006)	Man (2013)	Novakovic-Agopian (2018)	Chen (2011)	**TOTALS**
Internal validity											
Selection and allocation											
True randomisation of assignment of participants to groups?	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	**10/10**
Allocation to treatment groups concealed to researchers?	~	✓	✓	~	~	**X**	✓	**X**	✓	~	**4/10**
Treatment groups similar at baseline?	**X**	✓	**X**	✓	**X**	**X**	**X**	✓	✓	~	**4/10**
Administration of intervention/exposure											
Participants blind to treatment?	**X**	✓	**X**	**X**	**X**	**X**	✓	**X**	**X**	**X**	**2/10**
Those delivering treatment blind to treatment?	~	*NA*	~	~	~	~	**X**	~	**X**	**X**	**3/10**
Treatment groups treated identically other than intervention of interest?	✓	✓	✓	~	**X**	✓	✓	✓	✓	✓	**8/10**
Assessment, detection, and measurement of outcome											
Outcome assessors blind to treatment?	~	✓	✓	~	~	✓	✓	**X**	✓	~	**5/10**

Outcomes measured in same way for treatment groups?	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	**10/10**
Outcomes measured in reliable manner? (rater reliability)	~	~	~	~	~	~	~	~	✓	~	**1/10**
Outcomes measured in valid manner?(test validity)	**X**	✓	**X**	✓	✓	✓	✓	✓	✓	**X**	**7/10**
Outcomes measured in a reliable manner? (test reliability)	**X**	✓	**X**	✓	✓	✓	✓	✓	✓	**X**	**7/10**
Participant retention											
Was follow up complete, if not were differences between groups described and analysed?	**X**	✓	✓	✓	**X**	**X**	✓	**X**	✓	✓	**6/10**
Statistical conclusion validity											
Were participants analysed in the groups to which they were randomized?	**X**	✓	**X**	✓	✓	~	✓	**X**	**X**	✓	**5/10**
Was appropriate statistical analysis used?	**X**	**X**	**X**	**X**	**X**	**X**	**X**	✓	**X**	**X**	**1/10**
Was the trial design appropriate and any deviations from the standard RCT design accounted for?	**X**	**X**	**X**	**X**	**X**	**X**	✓	**X**	**X**	**X**	**1/10**

Note. ~ = unclear/not provided.

Participant Characteristics

Appendix I describes the characteristics of participants included within the reviewed studies. Fifty percent of the studies only included participants with TBI. Eleven out of 12 studies reported the age range of participants in some format. Ten studies reported the gender of participants. Within 60% of these studies, the majority of participants recruited were male. Eleven studies reported the length of time elapsed since injury. Time elapsed since injury varied widely between studies, ranging from two months to 14 years. Only seven out of 12 studies provided information on severity of injury, and most studies that reported severity of injury did not restrict severity to one category (e.g. mild/moderate/severe). The exception was the study by Fong and Howie (2009), who included participants with moderate ABI only.

Discussion

To the authors knowledge, this is the first systematic review focused on interventions aiming to improve cold executive functioning amongst adults with ABI. Interventions from 12 studies were reviewed.

Outcome of Interventions for Cold Executive Functioning

This literature review sought to assess the efficacy of interventions aimed at improving performance on cold executive functioning outcome measures amongst adults with ABI. In terms of efficacy, results showed that of the studies targeting 'executive function' only 50% evidenced an intervention-specific improvement. Of the studies targeting 'problem-solving', 60% evidenced an intervention specific improvement, whilst the remaining 40% demonstrated improvement over time regardless of the intervention participants received. All studies that aimed to improve more specific sub-constructs of cold executive functioning such as 'mental flexibility' and 'attention and executive function' evidenced an intervention-specific improvement. In terms of efficacy, this cautiously implies that targeted interventions aimed at improving a specific sub-construct of cold executive functioning (rather than executive functioning overall) may confer more benefit.

This literature review further sought to assess whether more efficacious interventions shared any characteristics. Of the eight studies that evidenced a significant intervention related improvement on cold executive function outcome measures, five studies included experimental interventions grounded in pre-existing neuropsychological theory which emphasised the role of attention in the maintenance of cold executive functioning (Chen et al., 2011; Fong & Howie, 2009; Novakovic-Agopian et al., 2018; Sohlberg et al., 2000; Tornås et al., 2016). This cautiously implies that this family of intervention may be more efficacious in promoting the maintenance of cold executive functioning.

Fong and Howie (2009) provided Metacomponential skills training to their experimental group to improve problem-solving abilities, guided by the theoretical concept that problem-solving can be broken down into primary meta-components: defining the problem, representing the problem, planning problem-solving strategies, monitoring selected strategies, and evaluating outcomes (Clements & Nastasi, 1990). The study further included everyday attention training as part of the intervention, guided by the hypothesis that individuals with ABI may oversimplify problems by failing to pay attention to salient information relating to the problem. This can render the meta-component of representing the problem challenging (Shallice & Burgess, 1991), and hinder overall problem-solving ability.

The studies by Chen et al. (2011), Novakovic-Agopian et al. (2018) and Tornås et al. (2016) all provided a variant of Goal Management Training ([GMT], Levine et al., [2011]), to their experimental groups. The training is based on a theory of sustained attention (Robertson & Garavan, 2000) which proposes that impairments of sustained attention may cascade to cause impairment in higher-order cognitive processes such as executive functioning.

Tornås et al. (2016) provided GMT (Levine et al., 2011) to their experimental group to improve executive functioning. The aim of the training was to teach participants to monitor and adjust their goal planning appropriately and consisted of five steps: orienting awareness to situation, defining the goal of the task, listing subgoals, learning the subgoals and evaluating whether action taken helped to move the goal forward. Both Novakovic-Agopian et al. (2018) and Chen et al. (2011) provided a variant of GMT, goal-oriented attentional self-regulation training ([GOALS], Novakovic-Agopian et al., 2007; 2011). The aim of the training was to improve attention and executive function, and mental flexibility. The GOALS intervention emphasized two components: the regulation of distractibility (e.g. redirection of attention), and the identification of personally relevant and feasible goals.

Sohlberg et al. (2000) provided Attention Process Training ([APT], Park et al., 1999; Sohlberg et al., 1994; Sohlberg & Mateer, 1987) to their experimental group to improve executive functioning. The intervention was grounded in the treatment model of Sohlberg and Mateer (1987), which specifies that attention can be split into five different components: focused, sustained, selective, alternating and divided. The theory posits that in order to recover overall attentional function, each of these five components must be rehabilitated in isolation. APT involved participants completing hierarchically organised tasks spanning five different components of attention: focused, sustained, selective, alternating and divided. The training aimed to place increasing demands on attentional and working memory mechanisms at a basic cognitive level, in order to rehabilitate high-order cognitive processes such as executive functioning ability.

It is important to note that none of the reviewed studies reported on the importance of the quality of therapeutic relationship between those providing the intervention and the study participants, despite extensive research to suggest that a strong therapeutic relationship between client and therapist is an important predictor of positive treatment outcome (Lambert & Barley, 2001; Steindl et al., 2023). This may be because neuropsychology as a discipline has traditionally overlooked this crucial contributor to engagement in rehabilitation.

Sustaining an ABI is likely to indicate that an individual has experienced a traumatic event. Many individuals who have sustained an ABI go on to present with symptoms that may be associated with post-traumatic stress disorder (PTSD), including experiences of intrusions (e.g. nightmares/flashbacks), avoidance (e.g. avoiding external reminders related to the trauma) and negative alternations in cognitions and mood (e.g. experiencing negative thoughts and difficulty experiencing positive emotions) (Iljazi et al., 20200).

The experience of a positive therapeutic relationship is crucial for individuals who have experienced any form of trauma, as such experiences pose a threat to physical and emotional safety. The therapeutic relationship can provide the foundation of a safe environment to rebuild a sense of security. In addition, many individuals who sustain an ABI and achieve 'physical recovery', often feel that their ongoing cognitive and emotional difficulties are invalidated as they cannot be 'seen' (Swift & Wilson, 2001). A positive therapeutic relationship allows for the acknowledgement of these difficulties to aid the process of engaging in rehabilitation. Experiencing a traumatic event can further result in heightened emotional arousal and difficulties with emotional regulation. A positive therapeutic relationship can ensure individuals feel able to express their emotions, as opposed to avoiding challenging situations (e.g. the process of rehabilitation) to ensure they 'stay in control' (Ardito & Rabellino, 2011). Future studies in this area may therefore benefit from considering the strength of the therapeutic relationship between study participants and intervention providers.

Outcome Measures used to Assess Cold Executive Functioning Improvements

A total of 25 outcome measures were used to assess improvements in cold executive functioning across the included studies. The majority of outcome measures were used in one study only, and no study used exactly the same outcome measures as another. The outcome measure used most frequently was the WCST. The WCST is sometimes regarded as the 'gold standard' neuropsychological test of EF (Kopp et al., 2019). Despite this, information regarding its core psychometric properties is lacking, as is research that assesses the convergent validity of the WCST with other tests of EF. This raises the possibility that EF has been operationalised by the tasks that are used to measure it, rather than an underlying theoretical construct (Cicerone, 2002).

There were some validity and reliability concerns with regard to the tests included in the study. First, certain authors used outcome measures with no accompanying validity and reliability data. For example, Soong et al. (2005) incorporated a self-efficacy checklist to measure change in problem-solving ability before and after intervention, however this checklist is not an established measure of problem-solving. If a test has not been established as valid or reliable, it provides very little information regarding change in the construct of interest.

Second, certain tests used by authors were established as adequately valid and reliable for measuring the construct of interest in certain populations but were used on a different participant population within the study. For instance, Jacoby et al. (2013) used the Executive Function Performance Test (EFPT) to measure changes in executive functioning before and after intervention. The EPFT is validated on 'psychiatric' and stroke populations, however not TBI populations, the focus of the study.

Third, certain tests used by authors were established as valid and reliable for the construct and population of interest but were then modified for the purpose of the study. For instance, Jacoby et al. (2013) used the Multiple Errands Test-Simplified Version (MET-SV) to measure change in executive functioning before and after intervention but modified the test to be administered in a large supermarket. Whilst the study notes that the content validity of the modified test was established by five occupational therapists, outcome measures ordinarily are required to undergo a lengthy and statistically driven process to establish validity and reliability.

The studies that best accounted for threats to validity and reliability were those that incorporated outcome measures established as valid and reliable for the construct and population of the interest. These studies also tended to use theoretically driven methodology to incorporate multiple different tests aimed to measure the construct of interest (e.g. Chen et al., 2011; Novakovic-Agopian et al., 2018). The benefit of using multiple and diverse tests to measure the construct of interest is that the convergent and concurrent validity of results is improved (Urbina, 2014).

The ecological validity of the included outcome measures within this study is likely to be low. This is because tests with higher external validity were often excluded by virtue of the fact that they were often linked to hot executive functions in some way, which was an exclusion criterion for this study. For example, the Dysexecutive questionnaire (DEX) from the BADS (Wilson et al., 1996) has high external validity because it asks participants to report on real life scenarios, but it was excluded from this study's analysis as it contained items pertaining to emotional regulation (e.g. participants are required to report on the likelihood of losing their temper). Very few tests used by authors showed adequate convergent validity. This is not a criticism of study authors but relates to the fact that there is no universally accepted definition of executive functioning.

It is important to highlight that neuropsychological tests are intended to measure cognitive functioning at a given time point and are not intended to act as outcome measures. This is because if neuropsychological tests are delivered at two time points, an individual will be

exposed to the same test material on retest, risking the emergence of residual and practice effects. Practice effects occur when an individual remembers some or all test items. Even if a parallel version of a test is used, practice effects can still occur because the 'process', as opposed to the content of the test can be learnt. It should therefore be considered that the improvement in cold executive functioning evidenced by many of the reviewed papers of this study may be the result of practice effects.

Level of Evidence and Quality of Studies

The level of evidence and results of the JBI appraisal checklists indicated that the quality of reviewed studies was generally poor. This limits the ability to accurately assess the efficacy of interventions aimed at improving performance on cold executive functioning outcome measures amongst adults with ABI.

The lack of participant and researcher blinding may have compromised the internal validity of the studies overall, as lack of double blinding may lead to an exaggeration of intervention effects and results that are biased in favour of the experimental group (Garattini et al., 2016). Baseline differences between treatment groups constituted a further threat to internal validity, as an imbalance in participant factors expected to affect treatment outcome (e.g. severity of ABI) can cause bias in the estimation of intervention effect (Jamieson, 1999).

The internal validity of studies was further compromised as a result of bias related to the assessment, detection and measurement of outcomes. Specifically, it is notable that just one of the 10 RCT studies showed that relevant outcome measures were rated by more than one assessor. Though performance on cold executive function outcome measures may appear objective in that results can be scored in accordance with clear marking criteria, there is still scope for error in that scores may be calculated incorrectly. The remaining nine studies may

have easily been improved through the inclusion of a further outcome assessor, which would have increased the reliability of outcome measurement (Barker et al., 2015).

The statistical conclusion validity across the 10 RCT studies was poor. Only half of these studies reported either no participant attrition or included ITT analyses. A lack of ITT analysis is problematic for several reasons. First, if participant attrition is not statistically accounted for, the effect of the intervention is likely to be inflated and a poor reflection of clinical practice. This is because the most likely 'real-life' clinical scenario is that some individuals will not complete an intervention or treatment. Second, participants who do not complete an intervention may do so on account of factors related to the intervention (e.g. finding it uncomfortable or unhelpful). It is important to understand and statistically account for these factors in order to ensure that interventions are as efficacious as possible. Third, the absence of an ITT analysis reduces the sample size of a study, which in turn reduces its statistical power (Gupta, 2011).

Small sample sizes were a major limitation of the studies included in this review, as were the lack of power analyses (carried out by only one study). Small sample sizes are problematic as they reduce statistical power, thus increasing the chance of a type II error being committed (VanVoorhis & Morgan, 2007). Small sample sizes were of particular concern for studies incorporating three experimental groups (e.g. Soong et al., 2005), as such studies require increased power to detect genuine effect. It should also be noted that several studies included in this review (e.g. Chen et al., 2011; Novakovic-Agopian et al., 2018) included a large number of outcome measures aiming to assess performance across a range of cognitive domains. This increases the chance of committing at type I error as a greater number of outcome measures increases the probability of an erroneously statistically significant result.

Several further factors limited the quality of the studies included in the review through rendering them difficult to replicate (Shrout & Rodgers, 2018). Specifically, there was a lack of reporting with regard to the duration, frequency, number and length of intervention sessions provided, and a lack of clarity regarding the job role and title of those providing the intervention (e.g. 'therapists' and 'examiners'). Studies not amenable to replication are problematic as the interventions they report on cannot be investigated through further confirmatory research, nor implemented and applied in clinical practice. For instance, most studies did not report on the level of training received by the individuals who delivered the interventions. It is therefore difficult to know whether the interventions reported on could be delivered in healthcare organisations such as the National Health Service (NHS).

Participant Characteristics

The reviewed studies reported a majority of male participants with ABI. This may be explained by evidence to suggest that men are more likely to sustain a TBI in work, combat and sports-related settings (Gupte et al., 2019). The reviewed studies included participants who sustained an ABI by a range of means, making comparability across studies challenging.

Not all studies reported on the severity (mild/moderate/severe) of ABI, and those which did set markedly different determination criteria. For instance, whilst some used baseline neuropsychological test scores to evidence degree of impairment and therefore injury severity, others used The Glasgow Coma Scale (Teasdale & Jennet, 1974), used to measure an individual's level of consciousness in the aftermath of brain injury. The severity of ABI has a huge impact on recovery trajectory (Barman et al., 2016; Ownsworth & Clare, 2006), and therefore should be accounted for in outcome analysis as it represents a significant confounding variable; research suggests that after one year, those who have sustained a mild TBI often demonstrate levels of functioning similar to non-injured individuals, whereas only 75% of

those with moderate TBI and 50% of those with severe TBI recover the ability to function independently over the same period (McCrea et al., 2021; Scholten et al., 2015).

Nearly all reviewed studies reported on time elapsed since participant injury. Time elapsed since injury is likely to have a significant influence on the effectiveness of interventions aimed at improving cold executive functioning, because the two-year-period after sustaining an ABI is regarded as the timeframe within which the majority of spontaneous recovery takes place (Fleminger et al., 2005; Holleman et al., 2018). It may therefore be anticipated that studies including participants at least 14 years post injury (e.g. Thaut et al., 2009), might report less cold executive functioning improvement compared to studies that included participants recruited in the direct aftermath of their ABI.

Despite good reporting, there was marked variation both within and between studies in terms of time elapsed since injury, with some participants within the same study having sustained an ABI years apart. This variation is problematic unless it is accounted for as a confounding variable in analysis because it is likely to have a significant effect on intervention effectiveness. Across reviewed studies there was limited information available regarding participant premorbid health. In addition, very few studies incorporated age as a covariate within statistical analyses. Accounting for this information may have been useful as previous research suggests that both age and premorbid health impact on recovery trajectory, and feasibly therefore response to intervention (Verdugo et al., 2019).

Clinical Implications

The results of this review suggest that the most effective interventions for cold executive functions focused on a specific sub-construct of executive functioning (e.g. ‘attention and executive function’ or ‘mental flexibility’). In addition, the most effective interventions for

cold executive functions in this review were grounded in neuropsychological theories which emphasised the role of attention in the maintenance of executive functioning.

Given the evidenced importance of attention in the maintenance of cold executive functioning, it may be useful to embed this family of intervention into clinical practice. This could be achieved via the development of clinical protocols guided by the following attention-based programmes: Metacomponential skills training (Fong & Howie, 2009), GMT (Levine et al., 2011), GOALS (Novakovic-Agopian et al., 2007; 2011), and APT (Park et al., 1999; Sohlberg et al., 1994; Sohlberg & Mateer, 1987). Such clinical protocols could enable a range of staff members (occupational therapists, physiotherapists, mental health nurses etc.) to deliver attention-based training within the NHS. This would ensure that ABI sufferers can access effective neuropsychological rehabilitation as soon as possible into their recovery journey (Hayashi et al., 2025).

From a clinical perspective, the lack of an accepted outcome measure for cold executive functioning may mean that clients often undertake a range of neuropsychological tests to evaluate possible impairments in this domain. Not only is this time-consuming and potentially anxiety-provoking, it increases the probability that clients will perform poorly on at least one measure of executive functioning, leading to a suspect deficit in executive functioning that may not exist. There is a clinical need to address the disproportionate number of tests used to measure cold executive functioning impairments, which may be achieved by further research.

Future Directions for Research

Future studies in this area may benefit from stronger methodology to increase their internal validity and statistical conclusion validity. This may be achieved via a double-blinding protocol. It may also be beneficial for studies to include more than one outcome rater, as this

represents a relatively straightforward way of increasing perceived and actual reliability of outcome measurement.

Further studies in this area may benefit from the inclusion of a more homogenous group of participants, in terms of severity of injury and time elapsed since injury. These characteristics should be clearly reported on, and where possible incorporated as confounding variables in statistical analyses. Homogeneity across participant groups increases the chance that intervention effects are due to the experimental effect of the intervention as opposed to random variation. The recruitment of homogenous participant groups may be supported by increased funding in ABI research, allowing collaborative multi-centre research efforts to support recruitment and the attainment of adequate sample sizes.

With regard to measuring cold executive functioning as a construct, the lack of an accepted universal measure means that different studies take markedly different approaches to cold executive functioning measurement, the result being that executive functioning research is rarely replicated and exceedingly difficult to compare. This is not a fault of study authors per se, but a reflection of the complexity of the construct. The results of this study suggest the most effective interventions for cold executive functioning impairments focus on a specific component of the construct. It logically follows that neuropsychological testing may also benefit from adopting this specific approach, 'narrowing' the construct of cold executive functioning into its constituent sub-constructs in order to reduce the number of tests purporting to measure a single construct.

Limitations

This review has a number of limitations. First, it was not within the scope of the review to analyse the association between severity and type of injury and the effectiveness of the intervention. Second, whilst the designation of hot and cold executive functions was based on

existing theory and research, the distinction between hot and cold executive functions is often not clear cut, with both working in tandem influence behaviour in 'real life' scenarios (Hongwanishkul et al., 2016). Third, many of the studies reviewed included a large number of outcome measures, and a decision was made as to which outcome measures captured change in cold executive functioning and therefore should be included in this review. Whilst these decisions were guided by the framework specified in the methods section, some decisions were, to some extent, subjective. Fourth, it was not possible to quantitatively synthesise the results of this systematic review in the form of a meta-analysis due to the heterogeneity of the outcome measures used to capture changes in cold executive functioning.

Conclusion

A systematic review was conducted on interventions aiming to improve cold executive functioning impairments amongst working age adults with ABI. The results suggested that the most efficacious interventions were targeted at a specific component of cold executive functioning and driven by a neuropsychological theory which accounted for the role of the cognitive process attention in cold executive functioning. The review incidentally found that 25 different outcome measures were used to measure changes in cold executive functioning. These findings supports the development of neuropsychological tests/outcome measures that target a specific component of cold executive functioning, to 'narrow' the complexity of the construct and decrease client burden.

References

Alderman, N., Burgess, P.W., Knight, C., & Herman, C. (2003). Ecological validity of a simplified version of the multiple errands shopping test. *Journal of the International Neuropsychological Society*, *9*(1), 31-44. https://doi.org/10.1017/s1355617703910046

Ardito, R. B., & Rabellino, D. (2011). Therapeutic alliance and outcome of psychotherapy: Historical excursus, measurements, and prospects for research. *Frontiers in Psychology, 2(*270). https://doi.org/10.3389/fpsyg.2011.00270

Barker, C., Pistrang, N., & Elliott, R. (2015). *Research methods in clinical psychology: An introduction for student and practitioners.* John Wiley & Sons.

Barker, T. H., Stone, J. C., Sears, K., Klugar, M., Tufanaru, C., Leonardi-Bee, J., Aromataris, E., & Munn, Z. (2023). The revised JBI critical appraisal tool for the assessment of risk of bias for randomized controlled trials. *JBI Evidence Synthesis*, *21*(3), 494-506. https://doi.org/10.11124/JBIES-22-00430

Barman, A., Chatterjee, A., & Bhide, R. (2016). Cognitive impairment and rehabilitation strategies after traumatic brain injury. *Indian Journal of Psychological Medicine*, *38*(3), 172-181. https://doi.org/10.4103/0253-7176.183086

Baum, C., Morrison, T., Hahn, M., & Edwards, D. (2007). *Executive Performance Test: Test Protocol Booklet.* Washington University.

Beecham, J., Perkins, M., Snell, T., & Knapp, M. (2009). Treatment paths and costs for young adults with acquired brain injury in the United Kingdom. *Brain Injury*, *23*(1), 30-38. https://doi.org/10.1080/02699050802590338

Bruns Jr, J., & Hauser, W. A. (2003). The epidemiology of traumatic brain injury: a review. *Epilepsia*, *44*(10), 2-10. https://doi.org/10.1046/j.1528-1157.44.s10.3.x

Burgess, P.W. (2010). Assessment of executive function. In J. Gurd, U. Kischka & J, Marshall (Eds.), *The handbook of clinical neuropsychology* (2nd ed., pp. 302-321). Oxford University Press.

Colantonio, A., Gerber, G., Bayley, M., Deber, R., Yin, J., & Kim, H. (2011). Differential profiles for patients with traumatic and non-traumatic brain injury. *Journal of Rehabilitation Medicine*, *43*(4), 311-315. https://doi.org/10.2340/16501977-0783

Centre for Disease Control and Prevention (2023). *Potential effects of a moderate or severe TBI.* https://www.cdc.gov/traumaticbraininjury/moderate-severe/potential-effects.html

Chan, R. C., Shum, D., Toulopoulou, T., & Chen, E. Y. (2008). Assessment of executive functions: review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology*, *23*(2), 201-216. https://doi.org/10.1016/j.acn.2007.08.010

Chavez-Arana, C., Catroppa, C., Carranza-Escárcega, E., Godfrey, C., Yáñez-Téllez, G., Prieto-Corona, B., De León, M.A., & Anderson, V. (2018). A systematic review of interventions for hot and cold executive functions in children and adolescents with acquired brain injury. *Journal of Paediatric Psychology*, *43*(8), 928942. https://doi.org/10.1093/jpepsy/jsy013

Chen, A. J. W., Novakovic-Agopian, T., Nycum, T. J., Song, S., Turner, G. R., Hills, N. K., Rome, S., Abrams, G.M., & D'Esposito, M. (2011). Training of goal-directed attention regulation enhances control over neural processing for individuals with brain injury. *Brain*, *134*(5), 1541-1554. https://doi.org/10.1093/brain/awr067

Cicerone, K.D. (2002). The enigma of executive functioning: theoretical contributions to therapeutic interventions. In P.J. Eslinger (Ed.), *Neuropsychological interventions: Clinical research and practice* (pp. 246-265). Guildford Press.

Clements, D. H., & Nastasi, B. K. (1990). Dynamic approach to measurement of children's metacomponential functioning. *Intelligence*, *14*(1), 109-125.

Coetzer, R., Carroll, E., & Ruddle, J. A. (2011). Depression, anxiety and employment status after traumatic brain injury. *Social Care and Neurodisability*, *2*(4), 200-207. https://doi.org/10.1108/20420911111188452

Colautti, L., Antonietti, A., & Iannello, P. (2022). Executive functions in decision making under ambiguity and risk in healthy adults: a scoping review adopting the hot and cold executive functions perspective. *Brain Sciences*, *12*(10), 1-22. https://doi.org/10.3390/brainsci12101335

Conners, C.K. (2000). *Connors' CPT II continuous Performance Test II.* Multi-Health Systems.

Daroff, R. B., & Aminoff, M. J. (Eds.). (2014). *Encyclopedia of the neurological sciences*. (2nd ed.). Academic Press.

De Luca, C. R., & Leventer, R. J. (2008). Developmental trajectories of executive functions across the lifespan. In V. Anderson, R. Jacobs, & P. J. Anderson (Eds.), *Executive functions and the frontal lobes: a lifespan perspective* (pp. 23–56). Taylor & Francis.

De Luca, R., Calabrò, R. S., & Bramanti, P. (2018). Cognitive rehabilitation after severe acquired brain injury: current evidence and future directions. *Neuropsychological Rehabilitation*, *28*(6), 879-898. https://doi.org/10.1080/09602011.2016.1211937

Delis, D.C., Kaplan, E., & Kramer, J.H. (2001). *Delis-Kaplan Executive Functioning System (D-KEFS).* The Psychological Corporation.

Department of Health (2017). *The adult social care outcomes framework 2017/2018 handbook of definitions.*

https://assets.publishing.service.gov.uk/media/5a82e915e5274a2e87dc3980/20171129_ASCOF_Handbook_of_Definitions_17-18.pdf

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135-168. https://doi.org/10.1146/annurev-psych-113011-143750

Downing, M., Hicks, A., Braaf, S., Myles, D., Gabbe, B., Cameron, P., Ameratunga, S., & Ponsford, J. (2021). Factors facilitating recovery following severe traumatic brain injury: a qualitative study. *Neuropsychological rehabilitation*, *31*(6), 889-913. https://doi.org/10.1080/09602011.2020.1744453J

Fleminger, S., & Ponsford, J. (2005). Long term outcome after traumatic brain injury. *British Medical Journal*, *331*(7530), 1419-1420. https://doi.org/10.1136/bmj.331.7530.1419

Fong, K. N., & Howie, D. R. (2009). Effects of an explicit problem-solving skills training program using a metacomponential approach for outpatients with acquired brain injury. *The American Journal of Occupational Therapy*, *63*(5), 525-534. https://doi.org/10.5014/ajot.63.5.525

Garattini, S., Jakobsen, J. C., Wetterslev, J., Bertelé, V., Banzi, R., Rath, A., Neugebauer, E.A.M., Laville, M., Masson, Y., Hivert, V., Eikermann, M., Aydin, B., Ngwabyt, S., Martinho, C., Geradi, C., Szmigielski, C.A.. Demotes-Mainard, J., & Gluud, C. (2016). Evidence-based clinical practice: overview of threats to the validity of evidence and how to minimise them. *European Journal of Internal Medicine*, *100*(*32*), 13-21.https://doi.org/10.1016/j.ejim.2016.03.020

Grant, D. A., & Berg, E. A. (1948). *Wisconsin Card Sorting Test* [Database record]. APA PsycTests.

Greig, T. C., Nicholls, S. S., Bryson, G. J., & Bell, M. D. (2004). The Vocational Cognitive Rating Scale: a scale for the assessment of cognitive functioning at work for clients with severe mental illness. *Journal of Vocational Rehabilitation*, *21*(2), 71-81.

Gupta, S. K. (2011). Intention-to-treat concept: a review. *Perspectives in Clinical Research*, *2*(3), 109-112. https://doi.org/10.4103/2229-3485.83221

Gupte, R. P., Brooks, W. M., Vukas, R. R., Pierce, J. D., & Harris, J. L. (2019). Sex differences in traumatic brain injury: what we know and what we should know. *Journal of Neurotrauma*, *36*(22), 3063-3091. https://doi.org/10.1089/neu.2018.6171

Hayashi, S., Kamo, T., & Momosaki, R. (2025). Effectiveness of early rehabilitation interventions in patients with traumatic brain injury using a large database. *Journal of Injury, Function, and Rehabilitation*, *17*(2), 170–177. https://doi.org/10.1002/pmrj.13243

Heaton, R.K., Miller, S.W., Taylor, M.J., & Grant, I. (2004). *Comprehensive norms for an expanded Halstead-Reitan Battery: demographically adjusted neuropsychological norms for African American and Caucasian adults (HRB).* Psychological Assessment Resources.

Hilbert M. (2020). Digital technology and social change: the digital transformation of society from a historical perspective. *Dialogues in Clinical Neuroscience*, *22*(2), 189–194. https://doi.org/10.31887/DCNS.2020.22.2/mhilbert

Holleman, M., Vink, M., Nijland, R., & Schmand, B. (2018). Effects of intensive neuropsychological rehabilitation for acquired brain injury. *Neuropsychological Rehabilitation*, *28*(4), 649-662. https://doi.org/10.1080/09602011.2016.1210013

Hongwanishkul, D., Happaney, K. R., Lee, W. S., & Zelazo, P. D. (2016). Assessment of hot and cool executive function in young children: age-related changes and individual differences. *Developmental Neuropsychology*, *28*(2), 617-644. https://doi.org/10.1207/s15326942dn2802_4

House of Commons. (2019). *Acquired brain injury*. (HC 2019/0111). https://researchbriefings.files.parliament.uk/documents/CDP-2019-0111/CDP-2019-0111.pdf

Iljazi, A., Ashina, H., Al-Khazali, H. M., Lipton, R. B., Ashina, M., Schytz, H. W., & Ashina, S. (2020). Post-traumatic stress disorder after traumatic brain injury: a systematic review and meta-analysis. *Neurological sciences*, *41*(10), 2737-2746. https://doi.org/10.1007/s10072-020-04458-7

Jacoby, M., Averbuch, S., Sacher, Y., Katz, N., Weiss, P. L., & Kizony, R. (2012). Effectiveness of executive functions training within a virtual supermarket for adults with traumatic brain injury: a pilot study. *Institute of Electrical and Electronics Engineers Transactions on Neural Systems and Rehabilitation Engineering*, *21*(2), 182-190. https://doi.org/10.1109/TNSRE.2012.2235184

Jamieson, J. (1999). Dealing with baseline differences: two principles and two dilemmas. *International Journal of Psychophysiology*, *31*(2), 155-161. https://doi.org/10.1016/S0167-8760(98)00048-8

Kendall, E., Shum, D., Halson, D., Bunning, S., & Teh, M. (1997). The assessment of social problem-solving ability following traumatic brain injury. *Journal of Head Trauma Rehabilitation, 12*(3), 68–78. https://doi.org/10.1097/00001199-199706000-00007

Kennedy, M. R., Coelho, C., Turkstra, L., Ylvisaker, M., Moore Sohlberg, M., Yorkston, K., Chiou H., & Kan, P. F. (2008). Intervention for executive functions after traumatic brain injury:

a systematic review, meta-analysis and clinical recommendations. *Neuropsychological Rehabilitation*, *18*(3), 257-299. https://doi.org/10.1080/09602010701748644

Kopp, B., Lange, F., & Steinke, A. (2021). The reliability of the Wisconsin card sorting test in clinical practice. *Assessment*, *28*(1), 248-263. https://doi.org/10.1177/107319111986625

Krasny-Pacini, A., Chevignard, M., & Evans, J. (2014). Goal management training for rehabilitation of executive functions: a systematic review of effectiveness in patients with acquired brain injury. *Disability and Rehabilitation*, *36*(2), 105-116. https://doi.org/10.3109/09638288.2013.777807

Lambert, M. J., & Barley, D. E. (2001). Research summary on the therapeutic relationship and psychotherapy outcome. *Psychotherapy: Theory, Research, Practice, Training, 38*(4), 357–361. https://doi.org/10.1037/0033-3204.38.4.357

Lawton, M. P., & Brody, E. M. (1969). Assessment of older people: self-maintaining and instrumental activities of daily living. *The Gerontologist, 9*(3), 179–186. https://doi.org/10.1093/geront/9.3_Part_1.179

Levine, B., Schweizer, T. A., O'Connor, C., Turner, G., Gillingham, S., Stuss, D. T., Manly, T., & Robertson, I. H. (2011). Rehabilitation of executive functioning in patients with frontal lobe brain damage with goal management training. *Frontiers in Human Neuroscience, (5),9, 1-9.* https://doi.org/10.3389/fnhum.2011.00009

Man, D. W. K., Poon, W. S., & Lam, C. (2013). The effectiveness of artificial intelligent 3-D virtual reality vocational problem-solving training in enhancing employment opportunities for people with traumatic brain injury. *Brain Injury*, *27*(9), 1016-1025. https://doi.org/10.3109/02699052.2013.794969

Manly, T., Hawkins, K., Evans, J., Woldt, K., & Robertson, I. H. (2002). Rehabilitation of executive function: facilitation of effective goal management on complex tasks using periodic auditory alerts. *Neuropsychologia*, *40*(3), 271-281. https://doi.org/10.1016/S0028-3932(01)00094-X

Marshall, R. C., Karow, C. M., Morelli, C. A., Iden, K. K., & Dixon, J. (2003). A clinical measure for the assessment of problem solving in brain-injured adults. *American Journal of Speech-Language Pathology, 12*(3), 333–348. https://doi.org/10.1044/1058-0360(2003/079)

Marshall, R., Karow, C., Morelli, C., Iden, K., Dixon, J., & Cranfill, T. (2004). Effects of interactive strategy modelling training on problem-solving by persons with traumatic brain injury. *Aphasiology*, *18*(8), 659-673. https://doi.org/10.1080/02687030444000237

McCrea, M. A., Giaimo, J. T., Barber, J., Temkin, N. R., Nelson, L. D., Levin, H. S., Dikmen, S., Stein, M., Bodien, Y.G., Boase, K., Taylor, S.R., Vassar, M., Mukherjee, P., Robertson, C., Diaz-Arrastia, R., Okonkwo, D.O., Markowitz, A.J., Manley. G.T., & TRACK-TBI Investigators. (2021). Functional outcomes over the first year after moderate to severe traumatic brain injury in the prospective, longitudinal TRACK-TBI study. *JAMA Neurology*, *78*(8), 982-992. https://doi.org/10.1001/jamaneurol.2021.2043

McDonald, B. C., Flashman, L. A., & Saykin, A. J. (2002). Executive dysfunction following traumatic brain injury: neural substrates and treatment strategies. *NeuroRehabilitation*, *17*(4), 333-344. https://doi.org/10.3233/NRE-2002-17407

Munn, Z., Barker, T. H., Moola, S., Tufanaru, C., Stern, C., McArthur, A., Stephenson, M. & Aromataris, E. (2020). Methodological quality of case series studies: an introduction to the JBI critical appraisal tool. *JBI Evidence Synthesis*, *18*(10), 2127-2133. https://doi.org/10.11124/JBISRIR-D-19-00099

Nastasi, B. K., & Clements, D. H. (1990). Metacomponential functioning in young children. *Intelligence*, *14*, 109-125. https://doi.org/10.1016/0160-2896(90)90017-N

Ng, B., Fong, K., Wong, R., Lau, A., Lee, C., Lau, G., et al. (1998). Validation of the Rivermead Behavioral Memory Test: Cantonese version. *Hong Kong Journal of Occupational Therapy, 9*(1), 18–24.

Novakovic-Agopian, T., Chen, A. J. W., Rome, S., Abrams, G., Castelli, H., Rossi, A., McKim, R., Hills, N., & D'Esposito, M. (2011). Rehabilitation of executive functioning with training in attention regulation applied to individually defined goals: a pilot study bridging theory, assessment, and treatment. *The Journal of Head Trauma Rehabilitation*, *26*(5), 325-338. https://doi.org/10.1097/HTR.0b013e3181f1ead2

Novakovic-Agopian, T., Chen, A.J.W., & Rome, S.R. (2007). *Goal Oriented Attentional Self-regulation (GOALS) training: instructor and participant manual.*

Novakovic-Agopian, T., Kornblith, E., Abrams, G., Burciaga-Rosales, J., Loya, F., D'Esposito, M., & Chen, A. J. (2018). Training in goal-oriented attention self-regulation improves executive functioning in veterans with chronic traumatic brain injury. *Journal of Neurotrauma*, *35*(23), 2784-2795. https://doi.org/10.1089/neu.2017.5529

Ownsworth, T., & Clare, L. (2006). The association between awareness deficits and rehabilitation outcome following acquired brain injury. *Clinical Psychology Review*, *26*(6), 783-795. https://doi.org/10.1016/j.cpr.2006.05.003

Oxford Centre for Evidence-Based Medicine (2009). *Levels of evidence.* https://www.cebm.ox.ac.uk/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *International Journal of Surgery, 88,* e105906. https://doi.org/10.1136/bmj.n71

Park, N. W., Proulx, G.-B., & Towers, W. M. (1999). Evaluation of the attention process training programme. *Neuropsychological Rehabilitation, 9*(2), 135-154. https://doi.org/10.1080/713755595

Patterson, T. L., Goldman, S., McKibbin, C. L., Hughs, T., & Jeste, D. V. (2001). UCSD Performance-Based Skills Assessment: development of a new measure of everyday functioning for severely mentally ill adults. *Schizophrenia Bulletin*, *27*(2), 235-245. https://doi.org/10.1093/oxfordjournals.schbul.a006870

Perna, R., Loughan, A. R., & Talka, K. (2012). Executive functioning and adaptive living skills after acquired brain injury. *Applied Neuropsychology*, *19*(4), 263-271. https://doi.org/10.1080/09084282.2012.670147

Peterson, E., & Welsh, M.C. (2014). The development of hot and cold executive functions in childhood and adolescence: are we getting warmer? In S. Goldstein & J. Naglieri (Eds.), *Handbook of executive functioning* (pp.45-65). Springer.

Pössl, J., Jürgensmeyer, S., Karlbauer, F., Wenz, C., & Goldenberg, G. (2001). Stability of employment after brain injury: a 7-year follow-up study. *Brain Injury*, *15*(1), 15-27. https://doi.org/10.1080/02699050118404

Rath, J. F., Simon, D., Langenbahn, D. M., Sherr, R. L., & Diller, L. (2003). Group treatment of problem-solving deficits in outpatients with traumatic brain injury: a randomised outcome study. *Neuropsychological Rehabilitation*, *13*(4), 461-488. https://doi.org/10.1080/09602010343000039

Raven, J., Raven, J. C., & Court, J. H. (2000*). Manual for Raven's progressive matrices and vocabulary scales*. The Psychological Corporation.

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills*, *8*(3), 271-276. https://doi.org/10.2466/pms.1958.8.3.271

Reitan, R.M., & Wolfson D. (1958). *The Halstead–Reitan Neuropsychological Test Battery: theory and clinical interpretation.* (2nd ed.). Neuropsychology Press

Robertson, I.H., & Garavan, H. (2000). Vigilant attention. In M.S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 631–640). MIT Press.

Salehinejad, M. A., Ghanavati, E., Rashid, M. H. A., & Nitsche, M. A. (2021). Hot and cold executive functions in the brain: a prefrontal-cingular network. *Brain and Neuroscience Advances*, *5*, 23982128211007769. https://doi.org/10.1177/23982128211007769

Scholten, A. C., Haagsma, J. A., Andriessen, T. M. J. C., Vos, P. E., Steyerberg, E. W., Van Beeck, E. F., & Polinder, S. (2015). Health-related quality of life after mild, moderate and severe traumatic brain injury: patterns and predictors of suboptimal functioning during the first year after injury. *Injury*, *46*(4), 616-624. https://doi.org/10.1016/j.injury.2014.10.064

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *298*(1089), 199-209. https://doi.org/10.1098/rstb.1982.0082

Shallice, T., & Burgess, P.W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain, 114*(2), 727-741. https://doi.org/10.1093/brain/114.2.727

Shames, J., Treger, I., Ring, H., & Giaquinto, S. (2007). Return to work following traumatic brain injury: trends and challenges. *Disability and Rehabilitation*, *29*(17), 1387-1395. https://doi.org/10.1080/09638280701315011

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: broadening perspectives from the replication crisis. *Annual Review of Psychology*, *69*, 487-510. https://doi.org/10.1146/annurev-psych-122216-011845

Sihvonen, A. J., Särkämö, T., Leo, V., Tervaniemi, M., Altenmüller, E., & Soinila, S. (2017). Music-based interventions in neurological rehabilitation. *The Lancet Neurology*, *16*(8), 648-660. https://doi.org/10.1016/S1474-4422(17)30168-0

Sohlberg, M., & Mateer, C. A. (1987). Effectiveness of an attention-training program. *Journal of Clinical and Experimental Neuropsychology*, *9*(2), 117–130. https://doi.org/10.1080/01688638708405352

Sohlberg, M., McLaughlin, K. A., Pavese, A., Heidrich, A., & Posner, M. I. (2000). Evaluation of attention process training and brain injury education in persons with acquired brain injury. *Journal of Clinical and Experimental Neuropsychology*, *22*(5), 656-676. https://doi.org/10.1076/1380-3395(200010)22:5;1-9;FT656

Sohlberg, M., Johnson, L., Paule, L., Raskin, S.A., & Mateer, C.A. (1994). *Attention Process Training II: a program to address attentional deficits for persons with mild cognitive dysfunction [rehabilitation materials].* Association for Neuropsychological Research & Development.

Soong, W., Tam, S. F., Man, W. K., & Hui-Chan, C. (2005). A pilot study on the effectiveness of tele-analogy-based problem-solving training for people with brain injuries. *International Journal of Rehabilitation Research*, *28*(4), 341-347. https://doi.org/10.1097/00004356-200512000-00006

Spivack, G., Platt, J. J., & Shure, M. B. (1976). *The problem-solving approach to adjustment: a guide to research and intervention.* Jossey-Bass.

Steindl, S. R., Matos, M., & Dimaggio, G. (2023). The interplay between therapeutic relationship and therapeutic technique: the whole is more than the sum of its parts. *Journal of Clinical Psychology*, *79*(7), 1686-1692. https://doi.org/10.1002/jclp.23519

Stergiou-Kita, M., Grigorovich, A., Damianakis, T., Le Dorze, G., David, C., Lemsky, C., & Hebert, D. (2017). The big sell: managing stigma and workplace discrimination following moderate to severe brain injury. *Work: Journal of Prevention, Assessment & Rehabilitation, 57*(2), 245-258. https://doi.org/10.3233/WOR-172556

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643. https://doi.org/10.1037/h0054651

Stuss, D. T., Stethem, L. L., & Pelchat, G. (1988). Three tests of attention and rapid information processing: an extension. *The Clinical Neuropsychologist*, *2*(3), 246-250. https://doi.org/10.1080/13854048808520107

Suchy, Y. (2009). Executive functioning: overview, assessment, and research issues for non-neuropsychologists. *Annals of Behavioral Medicine*, *37*(2), 106-116. https://doi.org/10.1007/s12160-009-9097-4

Swift, T. L., & Wilson, S. L. (2001). Misconceptions about brain injury among the general public and non-expert health professionals: an exploratory study. *Brain Injury*, *15*(2), 149-165. https://doi.org/10.1080/02699050117322

Teasdale, G., & Jennett, B. (1974). Assessment of coma and impaired consciousness: a practical scale. *The Lancet*, *304*(7872), 81-84. https://doi.org/10.1016/S0140-6736(74)91639-0

Thaut, M. H., Gardiner, J. C., Holmberg, D., Horwitz, J., Kent, L., Andrews, G., Donelan, B., & McIntosh, G. R. (2009). Neurologic music therapy improves executive function and emotional adjustment in traumatic brain injury rehabilitation. *Annals of the New York Academy of Sciences*, *1169*(1), 406-416. https://doi.org/10.1111/j.1749-6632.2009.04585.x

Tornås, S., Løvstad, M., Solbakk, A. K., Evans, J., Endestad, T., Hol, P. K., Schanke, A., & Stubberud, J. (2016). Rehabilitation of executive functions in patients with chronic acquired brain injury with goal management training, external cuing, and emotional regulation: a randomized controlled trial. *Journal of the International Neuropsychological Society*, *22*(4), 436-452. https://doi.org/10.1017/S1355617715001344

Tufanaru, C., Aromataris, E., Campbell, J., & Hopp L. (2017). Systematic reviews of effectiveness. In E. Aromataris & Z. Munn (Eds.), *Joanna Briggs Institute reviewer's manual* (pp. 3 – 10). The Joanna Briggs Institute.

Urbina, S. (2014). *Essentials of psychological testing* (2nd ed.). John Wiley & Sons.

VanVoorhis, C. W., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology*, *3*(2), 43-50. https://doi.org/10.20982/tqmp.03.2.p043

Verdugo, M. A., Fernández, M., Gómez, L. E., Amor, A. M., & Aza, A. (2019). Predictive factors of quality of life in acquired brain injury. *International Journal of Clinical and Health Psychology*, *19*(3), 189-197. https://doi.org/10.1016/j.ijchp.2019.06.004

Ward, J. (2020). *The student's guide to cognitive neuroscience.* Taylor & Francis.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale-III manual.* The Psychological Corporation.

Wesson, J., Clemson, L., Brodaty, H., & Reppermund, S. (2016). Estimating functional cognition in older adults using observational assessments of task performance in complex everyday activities: a systematic review and evaluation of measurement properties. *Neuroscience & Biobehavioral Reviews*, *100*(68), 335-360. https://doi.org/10.1016/j.neubiorev.2016.05.024

Williams, H., Caplan, B., Bogner, J., Brenner, L., Perrin, P. B., Niemeier, J. P., Mougeot, J., Vannoy, C.H., Hirsch, M.A., Watts, J.A., Rossman, W., Grafton, L.M., Guerrier, T.D., Pershad, R., Kingsbury, C.A., Bartel, S.W., & Whitney, M. P. (2015). Measures of injury severity and prediction of acute traumatic brain injury outcomes. *Journal of Head Trauma Rehabilitation*, *30*(2), 136-142. https://doi.org/10.1097/HTR.0000000000000026

Wilson, B. A., Alderman, N., Burgess, P. W., Emslie, H., & Evans, J. J. (1996). *BADS: Behavioural Assessment of the Dysexecutive Syndrome*. Pearson.

Yoshida, K., Ogawa, K., Mototani, T., Inagaki, Y., Sawamura, D., Ikoma, K., & Sakai, S. (2018). Flow experience enhances the effectiveness of attentional training: A pilot randomized controlled trial of patients with attention deficits after traumatic brain injury. *NeuroRehabilitation*, *43*(2), 183-193. https://doi.org/10.3233/NRE-172396

LUCY MCIVOR, BSC (Hons), MSc, MPhil

THE ASSESSMENT AND TREATMENT OF 'COLD' EXECUTIVE FUNCTIONING IMPAIRMENTS

Section B: Increasing the Clinical Utility of the SPANS-X Conceptual Flexibility Index: A Pilot study with Healthy Adult Participants

Word count: *7,926 (390)*

A thesis submitted in partial fulfilment of the requirements of
Canterbury Christ Church University for the degree of
Doctor of Clinical Psychology

APRIL 2024

SALOMONS INSTITUTE
CANTERBURY CHRIST CHURCH UNIVERSITY

Selected journal: Journal of Clinical and Experimental Neuropsychology

Abstract

Introduction: The Short Parallel Assessments of Neuropsychological Status-X (SPANS-X) is a neuropsychological test used for the assessment of acquired brain injury. Such assessment is crucial to ensure that clients receive appropriate intervention to enhance quality of life. However, the conceptual flexibility index (CFI) of the SPANS-X requires further development to improve its clinical utility. The aim of this pilot study was to improve the construct and convergent validity and internal consistency reliability of the CFI of the SPANS-X.

Methods: New theoretically-driven items were designed and added to the existing 3-and-1 concept test and similarities test of the CFI, based on the construct of conceptual flexibility. Four new subtests were also designed and added: associations, shape matrices, proverbs A and proverbs B. Fifty healthy participants completed the newly designed CFI online on Psytoolkit. Participant scores on the newly added items and subtests were then correlated with their scores on the original CFI to assess convergent validity, whilst Cronbach's alpha coefficients were used to evaluate internal consistency reliability. A multiple linear regression was also conducted to explore the impact of sex, age, education level, concussion, and neurodiversity on CFI score.

Results: Certain newly added items and subtests showed 'good' convergent validity with the original CFI (0.51). Certain newly added items and subtests increased the internal consistency reliability of the CFI from 'acceptable' to 'good (0.72 to 0.83). Age and education level had a significant impact on CFI score consistent with the original CFI findings, and suggesting these contextual variables are important to account for when interpreting client performance.

Conclusions: The CFI of the SPANS-X appears to increase its validity and reliability through the inclusion of new theoretically-driven items and subtests when tested with healthy participants, however more research and larger scale testing with clinical populations is needed.

Please see appendix J for selected research journal abstract guidelines.

Introduction

Impairments in ‘cold’ executive functioning are one of the most frustrating and inconvenient potential consequences of acquired brain injury (ABI) and can negatively impact an individual’s functional and vocational abilities (Chan et al., 2008; McDonald et al., 2002). Impairments may manifest as difficulties with starting or completing tasks, multi-tasking, thinking through problems, developing solutions, planning and organisation (Colautti et al., 2022; De Luca & Leventer, 2008).

Given such effects of cold executive functioning impairments, it is crucial that accurate neuropsychological tests exist to evaluate them. Such tests can assess for the presence or absence of impairment, delineate the pattern and severity of impairment, and clarify the cognitive effects of impairment on functional and vocational pursuits. This valuable information can then inform a neuropsychological treatment plan, with appropriate cognitive rehabilitation recommendations (Zucchella et al., 2018).

No unified theory of cold executive functioning exists, and as such, as a construct it is difficult to define and measure. Multiple different neuropsychological tests exist to assess cold executive functioning, including the Wisconsin Card Sort Test ([WCST], Grant & Berg, 1948), the Delis-Kaplan Executive Function System ([DKEFS], Delis et al., 2001), the Hayling and Brixton Tests (Burgess & Shallice, 1997), and the Behavioural Assessment of Dysexecutive Syndrome ([BADS], Wilson et al., 1996). The tests have different associated normative samples and interpretative metrics and correlate poorly (Manchester, 2004).

Pragmatically, the use of multiple different neuropsychological tests to assess for impairments in cold executive functioning is problematic. From a clinical perspective, clients may be required to complete a large and extensive range of neuropsychological tests as part of their assessments. In addition to being time-consuming and anxiety-provoking, this increases the

likelihood that clients will score at least one 'impaired' score simply by chance (Decker et al., 2012, Karr et al., 2017). This could lead to clinicians falsely suspecting an impairment that does not exist. Second, from a research perspective and as shown by the results of Section A, different studies aiming to evaluate effective interventions for cold executive functioning impairments take a varied and disparate approach to its measurement. Research is therefore rarely replicated and difficult to compare.

The dilemma of a potentially disproportionate number of neuropsychological tests to assess for impairments in cold executive functioning has several possible solutions. One solution is to incorporate the entirety of the multitude of tests that exist within this area into testing research, compare the results in terms of construct, convergent, and discriminant validity, to establish which tests come closest to measuring the 'true' construct of cold executive functioning (Pickens et al., 2010). This is pragmatically unrealistic, as it would require research participants to complete a vast battery of tests, introducing multiple confound errors in the process. It is also theoretically unrealistic, as cold executive functioning is a complex and multi-faceted construct unlikely to be captured by the results of one test alone (Baggetta & Alexander, 2016).

A more viable solution may be to 'narrow' the construct of cold executive functioning, to reduce the number of tests purporting to measure it. The results of Section A suggest that the most effective interventions for cold executive functioning impairments focus on a specific component of the construct. It logically follows that neuropsychological tests, as well as interventions, may benefit from this more specific and theory-led approach.

There are very few neuropsychological tests that focus explicitly on a specifically defined component of cold executive functioning. The Short Parallel Assessments of Neuropsychological Status-X ([SPANS-X], Burgess, 2021) has attempted this specific focus

via a Conceptual Flexibility Index (CFI). The SPANS-X has the further benefit of being brief to administer, which may reduce error due to client fatigue. Despite these benefits, at present, the CFI index of the SPANS-X is not performing optimally or at its presumed potential in terms of validity and reliability, which may limit its usefulness.

What Constitutes a Useful Neuropsychological Test?

Broadly speaking, a neuropsychological test may be classed as useful if it demonstrates adequate validity and reliability (Schoenberg & Scott, 2011). Test validity is the extent to which a test accurately corresponds to the construct it sets out or purports to measure, whilst test reliability is the extent to which a test performs in consistent, predictable ways (Barker et al., 2015). There are a range of validity and reliability subtypes.

Construct validity is the extent to which a test accurately measures all facets of a construct of interest (Strauss & Smith, 2009). Construct validity in its purest form can only be assessed indirectly, because the 'construct' is a latent (hidden) variable that cannot directly be observed (DeVellis & Thorpe, 2022). As an example, the executive functioning sub-construct of 'conceptual flexibility' is not an objective truth, but a theoretically-driven representation of a specific cognitive process that has potentially myriad forms of expression. Therefore, the design of a novel neuropsychological test for conceptual flexibility must rely heavily on theory to establish construct validity because, for hypothetical constructs, the only way to examine whether a test reflects construct validity is to examine whether the items conform to a theory (Smith, 2005).

Construct validity is estimated via multiple types of more operationally definable and measurable validities, e.g. convergent validity. Convergent validity is calculated by computing a correlation between the test being validated and another criterion that is believed to exemplify expression of the construct (Barker et al., 2015). Ordinarily, the other criterion is an established,

‘valid’ psychometric test that purports to measure the same construct. Convergent validity is considered ‘good’ if a correlation with a test purporting to measure the same construct is > 0.50 (Abma et al., 2016). The sensitivity and specificity of a neuropsychological test are further measures of its validity. Sensitivity refers to a test’s ability to identify the presence of a condition, whereas specificity refers to a test’s ability to identify the absence of a condition. If a neuropsychological test is not adequately sensitive or specific, it may fail to identify the presence of a condition in an individual with impairments, or incorrectly identify the presence of an impairment in a healthy individual (Akobeng, 2007).

Internal consistency is one type of reliability that reflects the homogeneity of items within a scale. A scale is internally consistent to the extent to which items are intercorrelated. Measurement theory supposes that items that strongly correlate with one another are logically connected because they therefore must also all have a strong correlation to the latent variable (i.e. the underlying construct) (DeVellis & Thorpe, 2022).

Internal consistency is associated with Cronbach’s (1951) coefficient alpha. Alpha is defined as the proportion of a scale’s total variance that is attributable to the latent variable underlying all items. All variation that is due to the latent variable is shared. Calculating alpha partitions the total variance amongst sets of items into ‘signal’ and ‘noise’ components. Alpha (the ‘signal’) = 1-error variance (the ‘noise’). Thus, the higher the alpha coefficient, the higher the proportion of variance that is purportedly attributable the latent variable and not error, and the higher the homogeneity of the scale (DeVellis & Thorpe, 2022). Alpha coefficients may be interpreted as follows; <0.50 is classed as unacceptable, 0.51-0.60 as poor, 0.71-0.80 as acceptable, 0.81-0.90 as good, and 0.91-0.95 as excellent (George & Mallery, 2003).

The Construct of Conceptual Flexibility

SPANS-X is unique in that it purports to measure the construct of 'conceptual flexibility', which may be considered a facet sub-construct of cold executive functioning. Conceptual flexibility is a SPANS-X specific term which was further operationalised for the purpose of this study. A healthy individual may demonstrate conceptual flexibility in everyday living through abstracting from the concrete to the generalisable, changing tact when the situation requires it, learning from feedback, and holding the 'bigger picture' in mind (Burgess, 2021).

The operationalisation of conceptual flexibility for the SPANS-X was theoretically-driven and derived from well-established and valid neuropsychological tests that measure similar constructs, such as 'cognitive flexibility' (Table 1). Conceptual flexibility can be thought of as similar to cognitive flexibility, except that conceptual flexibility has a narrower focus as it pertains to concepts specifically. A concept may be thought of as a changing characteristic taking a range of different forms, such as category, size, shape, use, colour, position, meaning and direction (Figure 1).

Table 1

Neuropsychological Tests used to Derive the Construct of Conceptual Flexibility

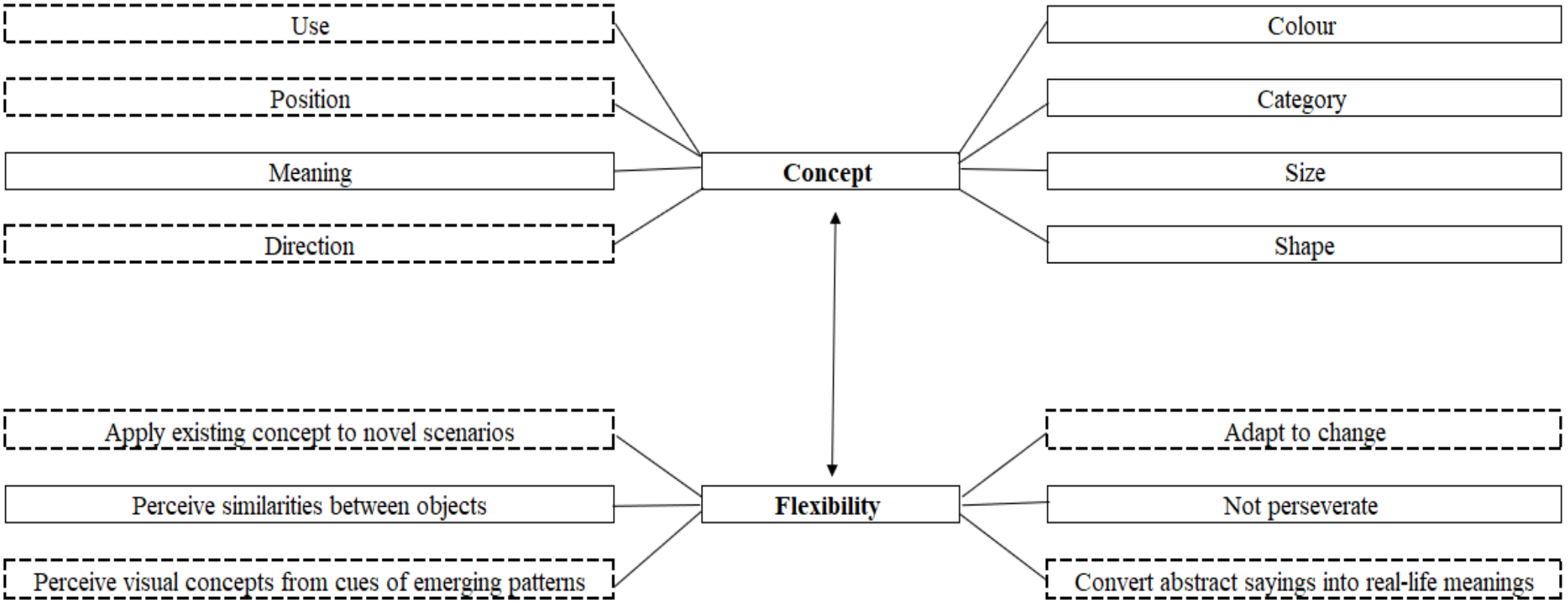
Test	Participant instructions	Relationship to construct of cognitive flexibility
Wisconsin Card Sort Test (Grant & Berg, 1948)	Participants are required to match stimulus cards based on certain characteristics (e.g. colour, shape, number). Participants are not given instructions on how to match the cards but are given feedback as to whether their choice is correct or incorrect.	Adapting to change/not perseverating Participants are required to adapt to change/not perseverate through shifting set when required. For instance, if participants are matching stimulus cards on the characteristic of colour for several correct trials, but then receive feedback that their colour match is incorrect, they are required to adapt to this change and shift set as quickly possible without perseverating on the previous characteristic (e.g. continuing to match by colour when this is no longer indicated).
Delis-Kaplan Executive Function System: Sorting Test (Delis et al., 2001)	Participants are provided with sets of stimulus cards, each of a different shape and with a single word printed in the centre. Each set has eight pairs of arrangement options: five have a perceptual concept sort and three have a verbal concept sort. In the first phase (free sorting) the participant is asked to sort the cards into equal groups and to describe the common concept of each group. In the second phase (structured) the examiner builds different arrangement variations, and the participant is required to identify and give an accurate description of each group.	Adapting to change/not perseverating Participants are required to adapt to change/not perseverate through shifting set when required. For instance, in the second phase, participants must name the concept sort that the examiner has used, and then must shift set when the examiner changes and uses a different concept to sort the stimulus cards.
Halstead-Retain Test Battery: Trail Making Test B (Reitan, 1958)	Participants are required to draw a line switching between numbers and letters in consecutive order (e.g. A to 1, B to 2, C to 3 etc.)	Adapting to change/not perseverating Participants are required to adapt to change/not perseverate through shifting set when required. In this instance, the 'set' is letters and numbers. If participants are unable to shift set, they will continue to join one set only (e.g. drawing a line from 1, to

		2, to 3, to 4, to 5) which indicates perseveration.
The Wechsler Adult Intelligence Scale III: **Perceptual Organization Index (Wechsler, 1997)**	The WAIS-III POI consists of the picture completion, block design and matrix reasoning subtests. Picture completion: Participants view a picture with an important part missing and are required to identify the missing part. Block design: Participants put together blocks in pattern to match a pattern shown to them. Matrix reasoning: Participants view an array of related pictures with one missing square and select a picture that fits this array from five options.	Perceiving visual objects from cues of emerging patterns Within all three subtests, participants are required to perceive visual objects from cues of missing elements and/or emerging patterns. Within the picture completion subtest, the emerging pattern is contained within one image with a missing element. Within the block design subtest, the emerging pattern is based on an established pattern that the participant must understand and mimic. Within the matrix reasoning test, the emerging pattern is spread across an array of related images.
The Wechsler Adult Intelligence Scale IV: Similarities Test (Wechsler, 2008)	Participants are required to state the similarity between two words.	Perceiving similarities between concepts In order to state the similarity between two words, participants must have skills related to verbal concept formation. An answer that identifies a difference as opposed to a similarity demonstrates a participant's inability to form concepts via creating a link between the two words. Scoring is further based on the abstractness of an answer, as the more capable a participant is of abstraction, the higher they will score on the test
Stanford-Binet Intelligence Scales: Verbal Analogies (Roid, 2003)	Participants are given two words that share a relationship with one another. Based on this relationship, participants are required to generate a fourth word that has the same relationship with the third word that the first two words have one another, for an example 'blue is to colour as two is to ____?'	Applying existing concepts to novel scenarios In order to apply existing concepts to novel scenarios, participants must first understand the conceptual relationship between the first two words (e.g. blue is one type of the colour category). They are then required to flexibly abstract the concept of blue being 'one type' of a category to a novel scenario, through considering what 'one type' of a category two may be (a number).

Delis-Kaplan Executive Function System: Proverb Test (Delis et al., 2001)	Participants must explain the real-life meaning behind the provided proverbs in the first condition of the test, then select the correct multiple-choice answer for the meaning of the proverb in the second condition.	Convert abstract sayings into real life meanings Proverbs are abstract expressions that convey messages about society. Successful interpretation of proverbs requires participants to understand the abstract meaning behind the proverb, rather than taking it as a concrete statement. An answer that indicates a participant has taken the problem 'literally' indicates a participant has been able to 'convert' the intended meaning into a real-life scenario.

Figure 1

Representation of the Construct of Conceptual Flexibility



Note. Boxes surrounded by a dashed border represent aspects of the construct of conceptual flexibility that are currently unexplored by the current SPANS-X CFI.

Validity and Reliability of the SPANS-X CFI

The SPANS-X CFI was designed to assess impairments in conceptual flexibility amongst ABI sufferers and therefore aims to maximise sensitivity and specificity for this population specifically. Therefore, the items are designed such that healthy populations who have not sustained an ABI are anticipated to score close to full marks (Burgess, 2021). A strength of the SPANS-X CFI is that it explicitly focuses on one component of cold executive functioning: conceptual flexibility. However, the CFI of the SPANS-X requires further development to remain clinically useful.

The current CFI of SPANS-X is derived from theoretical representations of conceptual flexibility (Figure 1), and consists of two subtests, the 3-and-1 concept test and the similarities test. The 3-and-1 concept test measures an ability to perceive more than one similarity between visual concepts, whilst the similarities test measures an ability to perceive similarities between verbal concepts (Burgess, 2021).

The current CFI of the SPANS-X may be considered a valid measure of the construct of conceptual flexibility, as it correlated highest with its closest theoretical equivalents; the Wechsler Adult Intelligence Scale III perceptual organization index (WAIS-III POI: , [Wechsler, 1997]) and the Trail Making Test B of the Halstead-Reitan Test Battery (HRTB, [Reitan, 1958]), with correlation values of 0.605 and -0.430 respectively. In addition, it correlated least with theoretically-dissimilar subtests: the WAIS-III working memory index (WMI) and the HRTB Trail Making Test A with correlation values of 0.251 and -0.260 respectively (Burgess, 2021). The current CFI of the SPANS-X may further be considered a valid measure of conceptual flexibility as a ‘good’ (0.51) correlation between participant scores exists on its two subtests, the 3-and-1 concept test and the similarities test. This correlation exits despite similarities being a verbal test and the 3-and-1 concept test being a visual test, characteristics which usually diverge within neuropsychological client groups due to right

hemisphere vs. left hemisphere lateralisation dominance and the differential effects brain injuries (Hall et al., 2008). What these two subtests appear to share, despite this contradiction, is the need to perceive concepts 'flexibly'.

Contrary to this promising evidence, the internal consistency reliability of the SPANS-X CFI as denoted by Cronbach's alpha is 0.72 ('acceptable'), the lowest of all the SPANS-X indices. This may imply that conceptual flexibility is not a valid construct, or that the current iteration of the CFI of the SPANS-X (which includes only two subtests) does not achieve sufficient statistical power to adequately detect evidence of the construct. The construct may be more comprehensively captured via the inclusion of additional, theoretically-driven items and subtests that seek to increase statistical power and 'tap into' aspects of the construct of conceptual flexibility that thus far remain unexplored, for instance the concepts of 'use', 'position' and 'direction', and aspects of flexibility such as applying existing concepts to novel scenarios, perceiving visual objects from cues of emerging patterns, adapting to change and converting abstract sayings into real-life meanings (Figure 1).

Contextual Interpretation of the SPANS-X CFI

Neuropsychological test development should aim to investigate any demographic or clinical variables that may plausibly affect an individual's overall test score (Salthouse, 1999). This is because an understanding of the impact of these variables will aid contextual interpretation of scores. The variables considered in this study were participant age, sex, gender, education level, presence/absence of concussion, and presence/absence of neurodiversity. The expected influence of these variables based on existing research are summarised below (Table 2).

Within this study, participants were also asked to report on the presence/absence of any neurological condition as the presence of a neurological condition may affect test scores. However, as expected, no participants met these criteria. This was primarily intended as a pilot

study aiming to improve the validity and reliability of the SPANS-X CFI. Part of ascertaining its validity in terms of sensitivity and specificity was to understand how healthy participants may be ‘expected’ to score. This step is a crucial precursor to eventually piloting the items with participants diagnosed with ABI.

Table 2

Anticipated Effects of Relevant Demographic and Clinical Variables on Expanded CFI Score

Variable	Rationale
Age	In line with the natural ageing process and subsequent 'on average' decline of visual acuity, motor skills and as yet undiagnosed cognitive impairment or dementia, a significant negative correlation will be observed between age and expanded CFI score (Burgess, 2021). From a theoretical perspective, age may decrease the ability to process novel stimuli and increase perseveration (Ridderinkhof et al., 2002).
Sex	There will be no significant group differences on expanded CFI score based on sex. The original CFI showed no group differences, and there is no anticipation a group difference will appear with the new subtests (Burgess, 2021).
Education level	There will be a significant positive correlation between education level and expanded CFI score, as demonstrated in the original CFI (Burgess, 2021). This is because those who are well educated may have had an enhanced opportunity to acquire skills that make them more successful when completing psychometric tests (e.g. Scheffels et al, 2023). This phenomenon has been coined 'test-wiseness', whereby educated individuals appear accustomed to the requirements of a testing situation, such as completing tasks with a time limit and the expectation of optimal effort. 'Test-wiseness' is primarily related to education opportunities as opposed to intelligence quotient (IQ),(Fernández, 2002; Marcopulos et al., 1997).
Presence/absence of concussion	To date, SPANS-X has been under-researched regarding its sensitivity and specificity for detecting impairment related to concussion (defined as being knocked on the head so hard that one has passed out). Based on existing research, it is hypothesised that there will be no significant group differences on expanded CFI score based on presence/absence of concussion, as deficits following concussion tend to relate to concentration difficulties as opposed to conceptual difficulties (Lovell, 1999; Willer & Leddy, 2006). If significant differences are found, it may be that the new CFI has a specificity level that is too low (e.g. it is incorrectly classing clients who have been concussed as having an ABI related conceptual flexibility impairment).
Presence/absence of neurodiversity	To date, SPANS-X has been under-researched regarding it sensitivity and specificity for detecting cognitive differences due to neurodiversity. Based on existing research, it is hypothesised that there will be no significant group differences on expanded CFI scores based on the presence/absence of neurodiversity. This is because whilst autistic individuals and individuals diagnosed with attention deficit hyperactivity disorder (ADHD) may score lower on neuropsychological tests that assess working memory and information processing speed, the CFI index of SPANS-X was not designed to load onto either of these constructs (Wilson, 2023). If significant differences are found, it may either be that the expanded CFI has a specificity level that is too low

	(e.g. it is incorrectly classing clients who identify as neurodiverse as having an ABI related conceptual flexibility impairment), that the expanded CFI is 'tapping into' constructs other than conceptual flexibility.

Note. Sex is reported on but not gender because within the study all 50 participants reported their gender as the same as their sex.

Research Aims

The current study aimed to improve the validity and reliability of the SPANS-X CFI via testing with a healthy participant population, whilst accounting for contextual influence of relevant participant variables on overall CFI score. Testing with a healthy participant population is a necessary precursor step towards increasing the clinical utility of the SPANS-X CFI for those with a clinical impairment.

An improved CFI would ultimately positively contribute to the development of neuropsychological tests that investigate a specific component of cold executive functioning (in this case, conceptual flexibility). This in turn may decrease the number of neuropsychological tests clients are required to undertake clinically, decreasing discomfort and anxiety in accordance with the NHS values of compassion and improving lives (NHS. 2024).

Hypotheses

1. The addition of new, theoretically-driven items to the CFI will improve its construct validity through more comprehensively capturing facets of the construct of conceptual flexibility.
2. The new items added to the CFI will, in combination, demonstrate 'good' convergent validity with the current CFI, with a correlation > 0.50.
3. The new items added to the CFI will, in combination, demonstrate 'good' internal consistency reliability with the current CFI, with a correlation > 0.81.

4. Relevant clinical and demographic characteristics will demonstrate either non-significant or significant associations with the newly expanded CFI score:

a) Age will show a significant negative correlation with overall CFI score.

b) Education level will show a significant positive correlation with overall CFI score.

c) There will be no significant group differences on overall CFI score based on sex.

d) There will be no significant group differences on overall CFI score based on presence/absence of concussion.

e) There will be no significant group differences on overall CFI score based on presence/absence of neurodiversity.

Method

Design

A non-experimental, cross-sectional survey design was used.

Consultation

Prior to the main study, two individuals were consulted regarding the language and readability of the consent form and the information sheet. They were also asked to complete the online neuropsychological test and share their opinion of it in terms of usability, clarity and difficulty (Appendix K). The feedback from these individuals was then incorporated into the final participant documents and neuropsychological test. Both consulted individuals were healthy to reflect the expected 'norm-like' characteristics of participants.

Recruitment and Participants

A quota sampling method was used to ensure a range of ages and education levels were included in the sample to reflect these demographic characteristics in the general population. The rationale for this was two-fold: first, both age and education levels can have a significant impact on neuropsychological test scores, and it is crucial that neuropsychological tests are universally applicable (not just to young people who are university educated). Second, the recruitment of a diverse sample of participants increases the likelihood of statistical variance amongst results, which is key to understanding how sensitive a test may be to participant differences. It was therefore decided that no more than 60% of the sample should have attended university, and no more than 50% of the sample should be under the age of 30. As these quotas were approached, the inclusion criteria of the study were changed. Participants were recruited via a range of social media sites (Appendix L).

Inclusion and Exclusion Criteria

The initial inclusion criteria were participants aged 18-90, able to understand English and capable of providing informed consent. Participants were also required to have access to a

computer to complete the neuropsychological test. Individuals of all neuropsychological abilities were eligible to take part. However, it was expected that participating individuals would have no neurological condition.

Measures

Measures included demographic and clinical questionnaires followed by an online neuropsychological test.

Demographic and Clinical Questionnaires

Demographic questions captured information regarding sex, ethnicity, education, age and UK acculturation (see Table 3 for explanation). Clinical questions asked participants to report on whether they had ever sustained a concussion, and whether they had ever been diagnosed with a neurological condition (captured by questions asking participants to report on whether they had been diagnosed with an ABI, cognitive impairment or dementia) or a neurodiversity condition by a health and social care professional (Appendix M). The rationale for collecting this information was to ascertain if these variables had a significant effect on neuropsychological test score.

Sample

The final sample consisted of 50 participants (76% female), aged 20-87 years (M=40.32, SD=16.80). 54% of participants were over age 30. 0% of participants completed less than secondary education, 16% of participants completed secondary education, 32% of participants completed some college or vocational training, and 52% completed a university degree. The majority of participants identified as White British (84%), had English as their first language (94%), had spent 50% or more of their life in a western English-speaking country (92%), had completed 50% or more of their education in a western English-speaking country (94%), had not sustained a concussion (76%), did not identify as neurodiverse (80%), and had not been diagnosed with a neurological condition (100%) (Table 3).

Table 3

Participant Demographic and Clinical Characteristics

Demographic Variables		
	N	**%**
Sex		
Female	38	76
Male	12	24
Ethnicity		
White British	42	84
White any other White background	3	6
Chinese other	2	4
Mixed White and Asian	1	2
Asian any other Asian background	1	2
White Irish	1	2
UK acculturation		
English first language?		
Yes	47	94
No	3	6
50% or more of life spent in western English-speaking country?		
Yes	46	92
No	4	8
50 % or more of education spent in western English-speaking country?		
Yes	47	94
No	3	6
Education		
Less than secondary education	0	0
Secondary education completed	8	16
Some college or vocational training	16	32
University degree	26	52

	Mean (SD)	Range
Age in years	40.32 (16.80)	20-87

Clinical Variables		
	N	**%**
Concussion?		
Yes	12	24
No	38	76
Neurodiversity?		
Yes	10	20
No	40	80
Neurological condition?		
Yes	0	0
No	50	100

Neuropsychological Test Questions

The neuropsychological test completed by participants was an expanded version of the original SPANS-X CFI. It was designed and coded by the author on Psytoolkit over the course of one year (May 2022-2023). Psytoolkit is an open-access psychological testing instrument that allows researchers to program and run psychological surveys and experiments (Stoet, 2010; 2017). The design of the expanded version of the original SPANS-X CFI consisted of several stages.

1. The current CFI of the SPANS-X, i.e., the original 3-and-1 concept test (eight items in total) and the original similarities test (six items in total) were digitalised
2. Additional items were designed and added to the original 3-and-1 concept test (eight items in total) and the original similarities test (five items in total)
3. Four further subtests were designed: associations (10 items in total), shape matrices (12 items in total), proverbs A (five items in total) and proverbs B (five items in total)

The design of the additional items and subtests was an iterative, theoretically-driven process. The new items and subtests were designed with the aim of capturing aspects of the construct of 'conceptual flexibility' not yet explored by the original CFI of the SPANS-X (Figure 1, Table 1). Table 4 outlines how the newly added items and subtests aimed to capture additional aspects of the conceptual flexibility construct.

Table 4

An Explanation of how Additional Subtests and Items Captured New Aspects of the Construct of Conceptual Flexibility

Test	Theoretical aspect of conceptual flexibility captured	Examples
3-and-1 concept (original)	Measures ability to perceive conceptual similarities between visual objects across several characteristics (category, size, shape, and colour), and the ability to generate <u>two</u> correct answers.	Examples: <u>Category</u>: e.g., three squares contain animals, whilst the fourth contains a human <u>Size</u>: e.g., three squares are larger, whilst the fourth is smaller <u>Shape</u>: e.g., three items are squares, whilst the fourth is a rectangle <u>Colour</u>: e.g., three items are green, whilst the fourth item is purple
	Measures ability to not perseverate, through the ability to generate <u>two</u> correct answers.	Perseveration relates to a proneness to maintain set or inability to shift from one specific, or conceptually general, set to another, i.e., the ability to perceive / select more than one correct answer.
3-and-1 concept (new)	Measures ability to perceive conceptual similarities of visual objects across additional / new dimensions (position, use).	Examples: <u>Position</u>: e.g., three items are on an 'upper' level, whilst the fourth item is on a 'lower' level <u>Use</u>: e.g., three can be used to travel on land, whilst the fourth can be used to travel on water
	Measures ability to adapt to change.	Which means: Instead of two correct answers, participants are asked to generate <u>three</u> correct answers (a rule change), then <u>four</u>, then as many as they can

		generate in a 30-second time limit (an introduced time constraint, with assigned number limit parameter removed).
Similarities (original and new)	Measures ability to perceive abstract similarities between concepts (meanings).	Example: How are a camel and a train alike? Participants are required to perceive that a camel and a train are similar because they are both modes of transport.
Associations	Measures ability to apply existing concepts to novel scenarios.	Example: Carpet is to house as grass is to…? Participants are required to comprehend that a carpet is something that lines the floor of a house. They are then required to abstract the concept of 'lining the floor' to a novel scenario, through considering what grass may 'line the floor' of.
Shape matrices	Perceive visual objects from cues of emerging patterns.	Example: Participants are required to complete suggested visual patterns, with cues or concepts such as, up, down, number, rotation, etc.
Proverbs A	Measures ability to convert abstract sayings into real-life meanings.	Example: Participants are required to convert the abstract concept that 'the grass is always greener on the other side of the fence' to a 'real-life' concept with a moral meaning, e.g. that other people's lives and situations can appear better than one's own.
Proverbs B	Measures ability to convert abstract sayings into real-life meanings.	Proverbs B operates similarly to proverbs A but is an easier test as instead of being required to generate their own meaning, participants are provided with multiple choice options.

Quality Assurance of Measures

A thorough explanation of the validity and reliability of the current CFI of the SPANS-X can be found in the introduction section. Is it not possible to comment on the validity and reliability of the expanded CFI of the SPANS-X, as determining its validity and reliability is the purpose of the current study (thus will addressed in the discussion section).

Ethics

Ethical approval was provided by a University Ethics committee (Appendix N) Feedback on study results was fed back to the ethics panel and participants (Appendices O & P). The research was conducted in line with the ethical principles of the British Psychological Society (British Psychological Society, 2021). All participants were informed as to how their data would be used and stored after participation. All participants were required to view the information sheet (Appendix Q) prior to starting the experiment. The information sheet included information on how to seek emotional support if needed should the experiment be experienced as distressing in any way. All participants were required to give informed written consent prior to starting the experiment, in accordance with the Declaration of Helsinki (Appendix R). Forced responses meant that participants could not continue with the experiment if informed consent was not provided.

Responses were anonymised by storing participant personal information provided for the prize draw (e.g., email addresses) separately to demographic, clinical and test data. This was achieved via assigning participants a random number generated code at the first opportunity.

Procedure

Data collection took place from June 2023 to January 2024. Participants were invited, via social media, to click on a link to take part in the study. Participants were required to read the information sheet and provide informed consent before proceeding with the study. Participants were then asked to provide their first name and the first letter of their surname and invited to

provide an email address for inclusion in the prize draw and if they wished to receive a copy of the study results. Participants had the chance to win one of 20 £15 vouchers. Participants then completed a demographic and clinical questionnaire before being directed to the online neuropsychological test, the expanded CFI of the SPANS-X. Feedback was provided on participants' progression through the test at the halfway mark and before the final subtest. On average participants took 30 minutes to complete the test. The fastest completion time was 16 minutes, and the slowest completion time was 50 minutes.

Planned Data Analysis

Preliminary data analyses were conducted to prepare the data, check underlying assumptions of the data in conjunction with the planned statistical tests, and report descriptive statistics (Field, 2013). All analyses were conducted using SPSS Version 29.

Power Analysis

An a priori power analysis was conducted for one-sided Spearman's correlation coefficient tests with a minimum power required of 0.8 and significance at the 0.05 level. This indicated a required sample size of 27. An a priori power analysis was conducted for a multiple linear regression model with a minimum power required of 0.08, five fixed predictors and significance at the 0.05 level. This indicated a required sample size of 45 (Appendix S).

Construct Validity

Construct validity can only be assessed indirectly and was therefore not measured statistically. To improve the construct validity of the SPANS-X CFI, additional items and subtests were added in accordance with the theory of conceptual flexibility (Table 4). As construct validity is not possible to measure statistically, convergent validity was used as a proxy.

Convergent Validity

Convergent validity was assessed by calculating Spearman correlation coefficients between the original CFI and the newly added items and subtests. A correlation of -1.0 indicates a strong

negative correlation, whilst a correlation of +1.0 indicates a strong positive correlation. Spearman as opposed to Pearson correlation coefficients were used as the data were not normally distributed, as participants tended to score at the top end of the scale. One-tailed Spearman correlation coefficient tests were conducted as there was an a priori hypothesis about the direction of effect, specifically that the newly added items and subtests were theoretically derived and therefore expected to correlate positively with the original CFI. This was a pilot study with a small sample size, and several larger scale studies will be required to establish a finalised expanded CFI for the SPANS-X. As such, corrections for multiple tests were not used, so as not to exclude potentially useful items at this early stage of testing.

For the new items added to the 3-and-1 concept test, participant scores on each conceptually similar set of items were added and correlated with participant overall score on the original 3-and-1 concept test. Items were classed as conceptually similar if they asked participants to name the same number of similarities and differences (i.e. two, three, four, and as many as possible in 30 seconds respectively).

For the new items added to the similarities test, each individual item was correlated with participant overall score on the original similarities test. The rationale for correlating items on an individual basis were that they were designed to increase in difficulty and could therefore be considered as conceptually different from one another.

For the new subtests added (associations, shape matrices, proverbs A and proverbs B), participants overall score on each test was correlated with their original 3-and-1 concept test score and original similarities test score. The rationale for correlating subtests as a whole as opposed to individual items is that each item was designed to be conceptually similar and of a similar difficulty level.

New items and subtests were considered to demonstrate some degree of convergent validity with the original CFI if the correlation was significant (i.e. $p < 0.05$). To gain an overall

convergent validity score, the scores of all new items and subtests with a significant correlation were combined and then correlated with the original SPANS-X CFI.

Internal Consistency Reliability

Items and subtests that showed convergent validity with the existing CFI of the SPANS-X consequently progressed to the internal reliability stage of testing. Items that did not demonstrate significant correlations were excluded. The internal consistency of each subtest was assessed by computing a Cronbach's alpha score. In addition, 'Cronbach's alpha if item deleted' was calculated for each individual item, to indicate which items within the subset may be lowering the reliability of the subtest as a whole. An overall Cronbach's alpha for the newly expanded CFI was also calculated. The newly expanded CFI consisted of all items from the original CFI combined with all new items and subtests which passed both validity and reliability checks.

Linear Regression Analysis

To explore the impact of participant demographic and clinical variables on CFI score, a multiple stepwise linear regression analysis was conducted. The underlying statistical assumptions for conducting a stepwise multiple linear regression analysis were checked and met by the data (Appendix T). The stepwise model included variables that increased the probability of F by at least 0.05, but excluded variables if they increased the probability of F by less than 0.1. The dependent variable within this analysis was the expanded CFI score. This expanded CFI score included all original SPANS-X CFI items, and any new items that passed both validity and reliability checks. The predictor variables included within the linear regression model were sex, age, education level, presence/absence of concussion and presence/absence of neurodiversity.

The ethnicity of participants was not included as a predictor variable as these data were reported for monitoring purposes and was not anticipated to have an effect on expanded CFI score. The

presence/absence of a neurological condition was not included as a predictor variable as no participants reported the presence of a neurological condition. Acculturation level was excluded as a predictor variable. This was because a very small number of participants answered ‘no’ to the acculturation questions, so the data collected violated the assumptions of the regression analysis on account of very unequal group sizes.

Missing Data

Fifty participants completed the neuropsychological test, whilst two participants started the test but did not complete it. Their data were therefore excluded. A further participant completed the test but experienced a computer malfunction, and contacted the author to inform them that they were unable to properly see the screen when answering test questions. Their data were therefore excluded. Total data completion was therefore 94.3%.

Results

Descriptive Statistics

Table 5 shows the descriptive statistics for participant scores on different items and subtests.

Table 5

Descriptive Statistics of Participant Scores on Different Items and Subtests

Test	Total item number	Maximum possible Score	Mean score	Median score	Mode score (N)	SD of score	Minimum observed score	Maximum observed score observed
3-and-concept (original, 1-8)	8	32	29.20	30	32 (17)	3.65	15	32
3-and-1 concept (9-10)	2	8	6.28	6	8 (23)	1.20	0	8
3-and-1 concept (11-12)	2	12	10.22	11.50	12 (25)	2.37	2	12
3-and-1 concept (13-14)	2	16	10.80	11	8 (22) 12 (22)	3.85	0	16
3-and-1 concept (15-16)	2	14[a]	6.04	6	6.00 (17)	2.78	0	14
3-and-1 concept (total, 1-16)	16	82	62.54	64	64 (12) 66 (12)	10.75	22	80
Similarities (original, 1-6)	6	12	10.92	11	12 (21)	1.65	2	12
Similarities 7	1	2	1.72	2	2 (38)	0.54	0	2
Similarities 8	1	2	1.46	2	2 (36)	0.61	0	2
Similarities	1	2	1.62	2	2 (33)	0.57	0	2

9								
Similarities 10	1	2	1.72	2	2 (40)	0.61	0	2
Similarities 11	1	2	1.92	2	2 (47)	0.34	0	2
Similarities (total, 1-11)	11	22	19.36	20	21 (18)	3.06	2	22
Associations	10	20	17.86	18	18 (13)	2.35	7	20
Shape matrices	12	12	11.00	12	12 (30)	2.08	0	12
Proverbs A	5	10	6.88	7	8 (13)	1.99	2	10
Proverbs B	5	5	4.82	5	5 (42)	0.44	3	5

[a]. This was a free response question, so theoretically participants could score any number of points. The highest point number scored by a participant, however, was 14.

Convergent Validity

Table 6 shows the Spearman correlation coefficients and associated significance levels between the original 3-and-1 concept test and similarities test and the newly added items and subtests. It further shows an overall convergent validity score for participant scores on the original CFI and their scores on newly added items and subtests that displayed convergent validity.

Table 6

Correlations between Original and New CFI Subtests

	3-and-1 concept (original)	**Similarities (original)**	**CFI (original)**
3-and-1 concept (9&10)	0.22 (0.06)		
3-and-1 concept (11&12)	0.07 (0.33)		
3-and-1 concept (13&14)	0.13 (0.18)		
3-and-1 concept (15&16)	0.39 (0.002)**		
Similarities 7		0.02 (0.45)	
Similarities 8		-.0.09 (0.26)	
Similarities 9		0.34 (0.007)**	
Similarities 10		0.40 (0.002)**	
Similarities 11		0.33 (0.01)*	
Similarities (original)	0.26 (0.04)*		
Associations	0.26 (0.03)*	0.26 (0.03)*	
Shapes	0.04 (0.40)	0.26 (0.04)*	
Proverbs A	0.15 (0.16)	0.46 (p<0.001)***	
Proverbs B	-0.02 (0.43)	0.12 (0.12)	
CFI (expanded)			0.51 (p<0.001)***

Note. $^{*}p < 0.05$. $^{**}p < 0.01$. ***$p < 0.001$.

Internal Consistency Reliability

Table 7 shows the total Cronbach's alpha for each subtest, and the 'Cronbach's alpha if item deleted' value for each item individually.

Table 7

Internal Consistency Reliability of the Expanded SPANS-X CFI

Test	Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Cronbach's alpha if item deleted	Total Cronbach's alpha (intraclass correlation)	95% confidence interval	
						Lower bound	Upper bound
3-and 1 concept					0.75	0.63	0.84
3-and-1 concept 1	38.04	37.30	0.47	0.72			
3-and-1 concept 2	38.32	38.88	0.35	0.73			
3-and-1 concept 3	37.88	40.35	0.40	0.73			
3-and-1 concept 4	37.68	40.79	0.57	0.72			
3-and-1 concept 5	37.72	39.76	0.65	0.72			
3-and-1 concept 6	37.84	40.46	0.43	0.73			
3-and-1 concept 7	37.92	42.03	0.23	0.74			
3-and-1 concept 8	37.56	43.64	0.36	0.74			
3-and-1 concept 9	38.24	36.10	0.62	0.70			
3-and-1 concept 10	38.52	38.50	0.33	0.74			
3-and-1 concept 15	39.32	35.86	0.34	0.74			
3-and-1 concept 16	37.68	31.32	0.45	0.74			

Similarities					0.79	0.69	0.87
Similarities 1	**14.18**	**7.04**	**0.00**	**0.80**			
Similarities 2	14.24	6.02	0.60	0.76			
Similarities 3	14.30	5.68	0.57	0.76			
Similarities 4	14.54	5.23	0.45	0.78			
Similarities 5	14.58	5.29	0.52	0.76			
Similarities 6	14.32	5.94	0.41	0.78			
Similarities 9	14.56	5.31	0.54	0.76			
Similarities 10	14.46	5.27	0.50	0.77			
Similarities 11	14.26	5.71	0.75	0.74			
Associations					0.65	0.48	0.7
Associations 1	16.20	4.12	0.39	0.60			
Associations 2	**16.28**	**4.94**	**0.12**	**0.67**			
Associations 3	15.86	5.51	0.00	0.65			
Associations 4	**15.92**	**5.50**	**-0.06**	**0.67**			
Associations 5	16.16	4.42	0.35	0.61			
Associations 6	16.20	4.57	0.26	0.63			
Associations 7	15.98	4.59	0.47	0.60			
Associations 8	16.08	4.28	0.37	0.60			
Associations 9	16.12	4.15	0.51	0.57			
Associations 10	15.94	4.30	0.64	0.56			
Shape matrices					0.86	0.80	0.91
Shape matrices 1	10.02	3.90	0.73	0.85			
Shape matrices 2	10.08	3.54	0.69	0.84			
Shape matrices 3	10.04	3.71	0.75	0.84			
Shape matrices 4	10.04	3.79	0.64	0.85			
Shape matrices 5	10.02	3.90	0.73	0.85			
Shape matrices 6	10.06	3.57	0.77	0.84			
Shape matrices 7	10.08	3.87	0.35	0.86			

Shape matrices 8	10.16	3.44	0.54	0.85			
Shape matrices 9	10.16	3.40	0.58	0.85			
Shape matrices 10	10.12	3.54	0.55	0.85			
Shape matrices 11	**10.10**	**3.89**	**0.29**	**0.87**			
Shape matrices 12	10.12	3.62	0.48	0.86			
Proverbs A					0.46	0.18	0.67
Proverbs A 1	5.28	2.74	0.31	0.36			
Proverbs A 2	5.46	2.46	0.28	0.38			
Proverbs A 3	5.44	2.82	0.26	0.39			
Proverbs A 4	5.06	3.16	0.28	0.40			
Proverbs A 5	**6.28**	**3.14**	**0.13**	**0.49**			
Expanded SPANS-X CFI					0.83	0.75	0.89

Note. Individual items that increase Cronbach's alpha item if item is deleted are in bold.

Linear Regression Analysis

Table 8 displays the descriptive statistics for SPANS-X CFI score according to the predictor variables of the regression model.

Table 8

SPANS-X CFI Score Descriptive Statistics according to the Predictor Variables of the Regression Model

	N	Mean (SD) Score	Minimum score	Maximum Score
Sex				
Female	38	89.45 (10.03)	52	105
Male	12	85.17 (11.78)	57	97
Education				
Less than secondary education	0	N/A	0	0
Secondary education completed	8	84.38 (10.78)	69	100
Some college or vocational training	16	84.56 (14.01)	52	105
University degree	26	92.04 (6.15)	78	102
Concussion?				
Yes	12	85.50 (9.53)	69	97
No	38	89.34 (10.76)	52	105
Neurodiversity?				
Yes	10	88.00 (6.46)	78	96
No	40	88.53 (11.37)	52	105

Note. Education levels are reported in categories to provide information on group differences; however, education was entered into the regression model as interval variable to replicate the method used with previously collected SPANS-X data. The interval variable age was also a predictor variable included within the regression analysis but is not displayed here as it is not possible to report group differences for this format of variable.

A stepwise multiple linear regression model was used to predict SPANS-X CFI score based on sex, age, education level, concussion status and neurodiversity status of participants. For the final model a significant regression equation was found including the variables age and education level ($F(2,47)=8.67$, $p < 0.001$, with an R^2 of 0.27). The variables sex, concussion status and neurodiversity were excluded by the model.

Table 9 demonstrates a significant effect of age and education level on CFI score. On average, as age increases, CFI Score decreases. On average, as education level increases, CFI score increases. Table 9 demonstrates no significant effect of sex, presence of concussion, or presence of neurodiversity on CFI score.

Table 9

The Effect of Predictor Variables on Overall CFI Score

		Unstandardized coefficients		**Standardized coefficients**		
	Beta In	***B***	**SE**	***B***	**T**	**Significance**
Age		-0.23	0.08	-0.41	-3.23	0.002
Education		3.80	1.77	0.27	2.15	0.04
Sex=female	0.12				0.94	0.35
Concussion=no	0.19				1.48	0.15
Neurodiversity=no	0.11				0.84	0.41

Note. Reference variables: neurodiversity = yes, concussion = yes, sex = male. The unstandardized coefficient '*B*' is indicative of the slope of the line between the predictor variable and the independent variable (e.g. for every one unit increase in age, there is -0.25 unit decrease in CFI score). The standardized coefficient '*B*' are measures of effect size with no units, so can be used to directly compare the magnitude of the effects of the different predictors.

Discussion

The SPANS-X CFI focuses on the assessment of the cold executive functioning sub-construct of conceptual flexibility. The aim of this pilot study was to improve the validity and reliability of the SPANS-X CFI through testing with healthy participant populations, via the addition of theoretically-derived items and subtests.

Validity

Newly Added Items to the 3-and-1 Concept Test

The original 3-and-1 concept test of the SPANS-X CFI presented participants with four items and asked them to name two ways that three of the items were the same and different from the fourth. The similarities/differences were centred around the concepts of category, size, shape and colour. The items newly added to the 3-and-1 concept test expanded upon this through asking participants to name an increasing number of similarities/differences (two ways, three ways, four ways, and as many ways as possible within 30 seconds).

The added items that asked participants to name two similarities/differences approached significant convergent validity with participants' scores on the original 3-and-1 concept test. These items followed a very similar design to the original 3-and-1 concept items. A correlation approaching significance is therefore promising and may benefit from further validity testing with larger sample sizes.

The added items that asked participants to name three similarities/differences did not demonstrate convergent validity with participants' scores on the original 3-and-1 concept test as expected. As most participants tended to score full marks across both sets of items, the lack of statistical correlation in this case indicates that the group of participants scoring full marks on the original 3-and-1 concept test were not the same group of participants scoring full marks on these newly added items, despite the intention that they assess the same conceptual

flexibility construct. Considering explanations for this difference, it is interesting to note that the newly added items presented participants with four words (e.g. cow, pig, rat and pear), and one of the similarities/differences that participants were required to identify was one related to meaning; specifically, that a cow, a pig and a rat are all animals, whereas a pear is not. It is therefore possible that these items were tapping into a construct other that conceptual flexibility, e.g. verbal comprehension, which requires understanding and reasoning using concepts framed in words (Sternberg & Powell, 1983).

The added items that asked participants to name four similarities/differences did not demonstrate convergent validity with participants' scores on the original 3-and-1 concept test as expected. In explanation, most participants tended to score close to full marks on the original 3-and-1 concept test (as indicated by the median and modal score values). This was not the case for these items, whereby the two modal scores showed that participants tended to obtain only 50% or 75% of the marks available. This finding could be interpreted in one of two ways.

First, it is possible that these items were assessing a construct other conceptual flexibility, as participants were required to identify that geometric shapes were located at different positions on the screen. The use of the 'position' concept was intended to expand the different expressions of conceptual flexibility explored by the test, however it is possible that this use of 'position' instead assessed participants' visuospatial ability, which is a construct distinct from conceptual flexibility and is a measure of an individual's ability to encode visual images and consider the relationship of visual images to space (Tverskey, 2010). Crucially, visuospatial abilities are subject to large individual differences in healthy adult populations, which may explain the variation in participant scores seen within this study (Mervis et al., 1999), thus explaining the lack of statistical correlation observed for these items.

Second, it is possible that the items in question did assess conceptual flexibility, however were 'too difficult' compared to other items, meaning that healthy adult participants with no ABI struggled to gain full marks. Ideally, this should not be the case for the expanded CFI of the SPANS-X, as the primary aim of the test is not to distinguish between the ability levels of healthy adults but to maximise its sensitivity and specificity through identifying those with clinical levels of impairment in the conceptual flexibility domain as a result of ABI (Burgess, 2021).

The items that asked participants to generate as many similarities/differences as possible within a 30-second time limit showed significant convergent validity with participants' scores on the original 3-and-1 concept test. These items were intended to expand the construct of conceptual flexibility as participants were required to describe the concept of 'use' for certain items (e.g. one similarity was that the objects displayed could be used for driving). In addition, and in line with previous well-established tests of flexibility such as the WCST, the DKEFS Sorting Test, and the HRTB Trail Making Test B, these items required participants to not perseverate via demonstrating an ability to shift set quickly (as the test was time limited), therefore also adapting to change. Tentatively, significant findings of convergent validity in this case lends credence to this study's conceptualisation of conceptual flexibility as a construct (Figure 2).

Newly Added Items to the Similarities Test

Participants' scores on the newly added items (9, 10 and, 11) all showed significant convergent validity with participants' score on the original test (1-6). The items that did not correlate with the original test were similarities 7 and similarities 8.

Similarities 7 showed a mean and modal participant score broadly in line with the other similarities items, which suggested that it was not the difficulty of the question leading to a lack of correlation. Instead, it must be considered that participants scoring higher on the

original similarities test were not the participants also scoring higher on this item. One explanation for this is that similarities 7 was assessing a construct other than that of conceptual flexibility, though it is not clear what construct this might constitute. A further explanation is that, as one of the words provided to participants was an object regularly used by tradespeople and within certain forms of employment, certain participants had a particular association with this word that led them to answer to the question differently as compared to participants less associated with this particular word. This may indicate that the question needs refining to apply 'equally' to a broad audience of participants.

Of all the newly added similarities items, participants gained the lowest mean score on similarities 8 (gaining 73% of the marks available), despite the modal score being full marks. In contrast, the mean score on the original similarities test (1-6) indicated that on average participants gained 91% of the marks available. It may therefore be the case that similarities 8 was too difficult for participants as compared to other items, despite the fact that the similarities questions were designed to increase in difficulty from item 7 to item 11. Items 'too difficult' for healthy participants are not useful as the primary aim of the test is not to distinguish between the ability levels of healthy adults but to maximise its sensitivity and specificity through identifying those with clinical levels of impairment in the conceptual flexibility domain as a result of ABI (Burgess, 2021). A further possibility is that that the marking criteria assigned to this item was too stringent.

Newly Added Subtests

The finding that participant scores on the original 3-and-1 concept test significantly correlated with participant scores on the original similarities test is expected and reassuring, as these tests constituted the original CFI of the SPANS-X and were designed to theoretically converge (Burgess, 2021).

The finding that the newly added associations test correlated with both the original 3-and-1 concept and similarities tests suggest that the associations test may be a very good fit for a potentially updated SPANS-X CFI. All three of these tests are concept driven and draw their theoretical basis from established neuropsychological tests that appear to assess capabilities related to 'flexibility'; the 3-and-1 concept test assesses the ability to adapt to change via shifting set and avoiding perseveration (a concept derived from the WCST, the sorting test of the DKEFS and the HRTB Trail Making Test B), the similarities test assesses the ability to perceive similarities between objects (derived from the WAIS-IV similarities test), and the associations test assesses the ability to apply existing concepts from novel scenarios (derived from the verbal analogies component of the Standford-Binet intelligence scales). The results of this study, which appear to show a three-way relationship between participant performance based on capabilities related to 'conceptual flexibility' specifically, tentatively lend support to the existence of conceptual flexibility as a valid construct (Figure 2, relationship denoted by italicized text).

The finding that participant scores on the shape matrices test correlated with the original similarities test but not the 3-and-1 concept test is an interesting one. This is because the theoretical basis for the shape matrices test was the WAIS-III POI, which requires participants to perceive visual patterns. Therefore, the 3-and-1 concept test and the shape matrices test are both visual, whilst similarities test is verbal. This suggests that the shared variance between the similarities test and the shape matrices test is not related to the visual aspect of the test. Observing the neuropsychological tests that these new items and subtests were derived from, both the WAIS-IV similarities test and the WAIS-III POI involve an element of detecting 'patterns', be it from a verbal or visual perspective. Going forward, an ability to detect 'patterns' between concepts may be considered a key part of the conceptual flexibility construct (Figure 2, relationship denoted by bold text).

The finding that participant scores on proverbs A correlated with participant scores on the original similarities test but not the 3-and-1 concept test is more expected as the two former are verbal measures. The proverbs A subtest was theoretically derived from the DKEFS proverbs test, which requires participant to convert abstract sayings into real-life meanings. Whilst this is a different capability to that required for the similarities test (an ability to perceive similarities between concepts, which requires verbal concept formation), one possible underlying capability between both these subtests may be verbal concept comprehension (Figure 2, relationship denoted by purple outline).

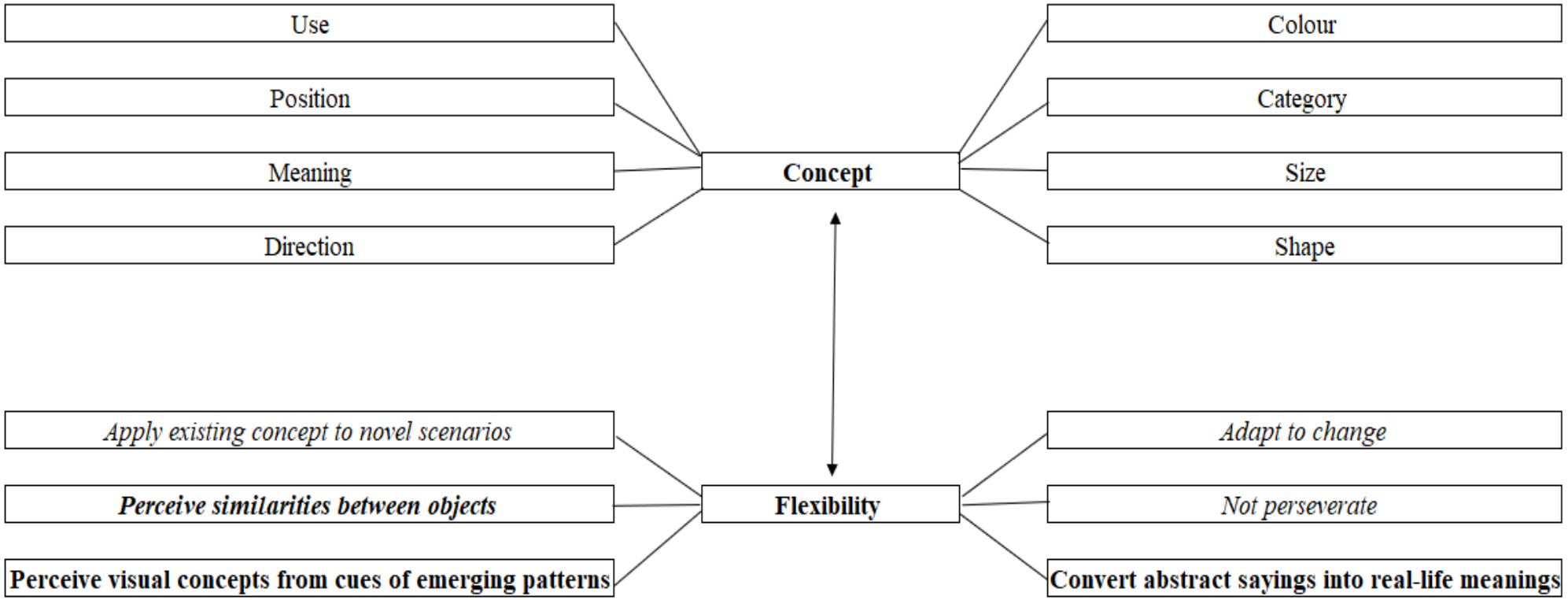
Participant scores on the proverbs B test did not show a significant correlation with either the original 3-and-1 concept test or the original similarities test. The mean and modal participant scores for the proverbs B test indicate that the vast majority of participants correctly answered every question, which was not the case for the other tests. This lack of variation in participant score may explain the lack of significant correlation. Based on this finding, the Proverbs B test should be excluded from further testing research, though this decision prompts an interesting debate regarding the threshold for the difficulty of the items included in the expanded version of the CFI. On the one hand, if items are too 'easy' for healthy participants (as appears to be the case with Proverbs B), then it is feasible that both healthy individuals and individuals with ABI are likely to score close to full marks, meaning that the items are unlikely to differentiate between those with and without a conceptual flexibility impairment. On the other hand, it is not possible to state definitively how individuals with ABI would score on proverbs B without further research and testing, and there is therefore a risk of excluding a subtest that may have proven potentially useful.

A further explanation for the lack of statistical correlation observed for the proverbs B subtests is that the small number of participants who did not score full marks on this test performed comparatively better on other tests, which again indicates the test may 'tap into' an underlying

construct other than conceptual flexibility. Crucially, proverbs B was at the end of the test battery which raises the possibility that participants' attentional abilities and motivation levels may have contributed (Mesulam, 2000).

Figure 2

Possible Capabilities underlying the Construct of Conceptual Flexibility



Note. Aspects of the construct are italicized and/or denoted in bold to represent relationships between constructs.

Reliability

Within the associations test, associations 2 and 4 decreased the reliability of the overall scale, and within the shapes matrices test, shapes matrices 11 decreased the reliability of the overall scale. These items may therefore benefit from being removed from the subscales to increase Cronbach's alpha. For the proverbs A test, proverbs A 5 decreased the reliability of the overall scale, and results showed that without this item, participants' mean overall score on the proverbs A test would have increased. Proverbs A 5 required participants to provide a meaning for the proverb 'only a fool tests the depth of a river with both feet.' This item was included as it is not considered a culturally common English proverb, so therefore may 'test' participants' conceptual flexibility (via their ability to convert abstract sayings into real-life meanings) more so than widely known proverbs. Participants may have struggled more to interpret the meaning of this proverb as a result. From a psychometric perspective, this item should therefore be excluded from the expanded CFI of the SPANS-X, however the potential negative consequences of this exclusion are that participants who have lower levels of acculturation are disadvantaged, as in some sense this item may have 'levelled the playing field' as it removed the advantage of prior cultural knowledge. The majority of participants in this study demonstrated a high level of acculturation so this hypothesis would require further research with a different participant group.

The results showed that when the items that increased the Cronbach's alpha of their subscales were removed, all items on the newly expanded SPANS-X CFI combined showed 'good' internal consistency reliability (George & Mallery, 2003).

Interpretability

The results of this study showed that sex was not a significant predictor of participant score on the expanded CFI of the SPANS-X, which was anticipated and replicates findings of the original CFI (Burgess, 2021). The results of this study showed that age was a significant predictor of score, with score decreasing with age. This replicates the findings of the original SPANS-X CFI (Burgess, 2021), and a wealth of research to suggest that, in line with the natural ageing process, mean neuropsychological mean test scores tend to decrease (and standard deviation widen) with age amongst healthy adult participants (Scheffels et al., 2023). The results further showed that participant' education level was a significant predictor of score, specifically that scores increased with higher education level. This also replicated the findings of the original SPANS-X CFI (Burgess, 2021) and further research which demonstrates that, across neuropsychological tests, higher education levels predict better test scores (Scheffels et al., 2023).

The results further showed that neurodiversity was not a significant predictor of score. This was anticipated given that whilst research shows that autistic individuals and individuals diagnosed with ADHD may score slightly lower on tests that assess working memory and information processing speed, the CFI of the SPANS-X was not designed to 'load' onto either of these constructs (Wilson, 2023). Moreover, the presence/absence of concussion was not a significant predictor of score. This was anticipated as extant research shows that having sustained a concussion does not ordinarily influence neuropsychological test scores and tends to cause concentration difficulties and visual disturbances as opposed to changes in conceptual flexibility (Lovell, 1999; Willer & Leddy, 2006). From a validity perspective these findings are positive, and indicate that the specificity level of the expanded CFI of the SPANS-X is appropriate (e.g. it does not differentiate between healthy participants and healthy participants

who identify as neurodiverse or have been concussed). This is crucial as the SPANS-X CFI is designed to assess for conceptual flexibility impairments in ABI specifically.

Overall, the results of the linear regression analysis indicate that when interpreting results of the expanded CFI, age and education level of individuals should be considered as contextual factors affecting test score. The current SPANS-X has age but not education related norms. The designation of education-related norms could therefore be a target for further research.

Clinical Implications and Future Research

Impairments in cold executive functioning are common following ABI. As such, it is crucial that accurate neuropsychological tests exist to evaluate these impairments to ensure effective neuropsychological treatment plans with appropriate cognitive rehabilitation goals. The SPANS-X CFI (Burgess, 2021) is designed to evaluate impairments in conceptual flexibility following ABI, however, requires development to improve its clinical utility.

The findings of this study suggest that the inclusion of certain new items and subtests may increase the validity and the reliability of the SPANS-X CFI index to a 'good' standard, as demonstrated by an overall convergent validity value of > 0.50, and overall Cronbach's alpha values of > 0.81. The findings further suggest that other added items and subtests decreased the validity or the reliability of the expanded SPANS-X CFI and should therefore not be brought forward to further testing phases. This included items and subtests that demonstrated too much response variation. This was because they 'loaded' onto theoretical constructs other than conceptual flexibility, such as visuospatial functioning or verbal reasoning (abilities which vary widely amongst healthy participants without implying any kind of deficit or impairment), or because they did not meet the appropriate difficulty level or were poorly worded or scored by the author.

The ultimate aim of the SPANS-X CFI must be to accurately assess impairments in conceptual flexibility, to facilitate quick and effective treatment for ABI sufferers. As demonstrated by the results of Section A of this study, attention-based programmes may represent the most efficacious treatments for impairments in conceptual flexibility. These programmes include Metacomponential skills training (Fong & Howie, 2009), GMT (Levine et al., 2011), GOALS (Novakovic-Agopian et al., 2007; 2011) and APT (Park et al., 1999; Sohlberg et al., 1994; Solhberg & Mateer, 1987). However, in order to progress towards the development of these attention-based intervention programmes, further refinement of the SPANS-X CFI is needed. In the future, it will be useful to calculate the convergent and divergent validity of the SPANS-X CFI with related neuropsychological measures, in addition to gaining qualitative feedback on individual SPANS-X CFI test items.

Limitations

The use of the alpha coefficient may be considered a limitation of this study. Alpha coefficient scores that have a lower number of items associated with them tend to have lower reliability, whereas scores that have a higher number of items associated with them tend to have higher reliability. Therefore, a higher alpha correlation coefficient may be considered a direct result of the additional items added. A further criticism of alpha is that reporting a single estimate of alpha without a confidence interval does not convey the observable variability in the data, leading to false confidence in the result (DeVellis & Thorpe, 2022).

The alpha coefficient was felt to be an appropriate measure of reliability within this study for the following reasons. First, alpha is a conservative estimate, and it often underestimates the lower bound of reliability. This reduces the risk of inflated results. Second, to overcome certain shortcomings associated with the alpha coefficient, a confidence interval (95%) for alpha was determined. Third, it is important to note that this study aimed to improve the CFI index of the SPANS-X. All internal consistency calculations for the SPANS-X have thus far been calculated

with alpha (Burgess, 2021). To allow for direct comparability, an alpha coefficient was also used for this study.

The absence of a factor analysis may also be considered a limitation of this study. Factor analysis allows complex sets of variables to be simplified using statistical analysis to explore the underlying common variance that explains the relationships between said variables (Tavakol & Wetzel, 2020). Factor analysis was not used within this study for two reasons. First, general guidelines for adequate participant sample size in factor analysis is as follows; 100= poor, 200= fair, 300= good, 500= very good (Comrey & Lee, 1992). The sample size of this small-scale study was 50, which does not approach the bounds for even a poor factor analysis. The use of factor analysis in the case may therefore have led to meaningless or misleading results. The sample size of 50 was the result of an extended recruitment effort, as the study took at least half an hour to complete. Second, this was a pilot study that aimed to develop a theory of conceptual flexibility whilst experimenting with potential items and subtests as a precursor to a larger scale study.

This pilot study involved testing with a healthy participant population, which is an essential precursor to improving the clinical utility of the SPANS-X CFI in the long term. However, it will be important to repeat the neuropsychological test designed for this study with a wider sample of participants, including those who have sustained an ABI or other neurological conditions. This is crucial to understand if and how healthy populations and ABI populations may score differently on the test.

Conclusion

There is evidence to suggest that the most effective neuropsychological tests for cold executive functioning impairment may focus on a specific component of the construct. The SPANS-X has significant potential as it focuses on conceptual flexibility specifically. This pilot study aimed to improve the validity and reliability of the SPANS-X CFI via testing with a healthy participant population, as a precursor to testing with clinical populations in the future. New theoretically-driven items and subtests were designed and added to the existing CFI, based on the construct of conceptual flexibility. Certain newly added items and subtests showed 'good' convergent validity and internal consistency reliability. Age and education level had a significant impact on CFI score, suggesting these contextual variables are important to account for when interpreting test scores.

References

Abma, I. L., Rovers, M., & van der Wees, P. J. (2016). Appraising convergent validity of patient-reported outcome measures in systematic reviews: Constructing hypotheses and interpreting outcomes. *BMC Research Notes*, *9*(1), 1-5. https://doi.org/10.1186/s13104-016-2034-2

Akobeng, A. K. (2007). Understanding diagnostic tests 1: sensitivity, specificity and predictive values. *Acta Paediatric*, *96*(3), 338-341. https://doi.org/10.1111/j.1651-2227.2006.00180.x

Baggetta, P., & Alexander, P. A. (2016). Conceptualization and operationalization of executive function. *Mind, Brain, and Education*, *10*(1), 10-33. https://doi.org/10.1111/mbe.12100

Barker, C., Pistrang, N., & Elliot, R. (2015). *Research methods in clinical psychology: An introduction for student and practitioners* (3rd ed.). Wiley Blackwell.

British Psychological Society (2021). *Code of ethics and conduct.* http://www.bps.org.uk/node/1714

Burgess, G.H. (2021). *SPANS-X: Short Parallel Assessment of Neuropsychological Status-Extended.* Hogrefe Ltd.

Burgess, P. W., & Shallice, T. (1997). *The Hayling and Brixton Tests.* Thames Valley Test Company.

Chan, R. C., Shum, D., Toulopoulou, T., & Chen, E. Y. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology*, *23*(2), 201-216. https://doi.org/10.1016/j.acn.2007.08.010

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Colautti, L., Antonietti, A., & Iannello, P. (2022). Executive functions in decision making under ambiguity and risk in healthy adults: A scoping review adopting the hot and cold executive functions perspective. *Brain Sciences*, *12*(10), 1-22. https://doi.org/10.3390/brainsci12101335

Comrey, A.L., & Lee, H.B. (1992). *A first course in factor analysis* (2nd ed.). Lawrence Erlbaum Associates, Inc.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297–334. https://doi.org/10.1007/BF02310555

De Luca, C. R., & Leventer, R. J. (2008). Developmental trajectories of executive functions across the lifespan. In V. Anderson, R. Jacobs, & P. J. Anderson (Eds.), *Executive functions and the frontal lobes: A lifespan perspective* (pp. 23–56). Taylor & Francis.

Decker, S. L., Schneider, W. J., & Hale, J. B. (2012). Estimating base rates of impairment in neuropsychological test batteries: A comparison of quantitative models. *Archives of Clinical Neuropsychology*, *27*(1), 69-84. https://doi.org/10.1093/arclin/acr088

Delis, D.C., Kaplan, E., & Kramer, J.H. (2001). *Delis-Kaplan Executive Functioning System (D-KEFS).* The Psychological Corporation.

DeVellis, R.F., & Thorpe, C.T. (2022). *Scale development: Theory and applications* (5th ed.). SAGE Publications.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135-168. https://doi.org/10.1146/annurev-psych-113011-143750

Fernández, A. L. (2022). Education, the most powerful cultural variable? In A. L. Fernández & J. Evans (Eds.), *Understanding cross-cultural neuropsychology: Science, testing, and challenges* (pp. 44–58). Routledge.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (3rd ed.). SAGE Publications.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference.* 11.0 update (4th ed.). Allyn & Bacon.

Grant, D. A., & Berg, E. A. (1948). *Wisconsin Card Sorting Test* [Database record]. APA PsycTests.

Hall, J.J., Neal, T.J., & Dean, R.S. (2008). Lateralization of cerebral functions. In A.M. Horton, Jr. & D. Wedding (Eds.), *The neuropsychology handbook* (3rd ed., pp. 183-214). Springer Publishing Company.

Karr, J. E., Garcia-Barrera, M. A., Holdnack, J. A., & Iverson, G. L. (2017). Using multivariate base rates to interpret low scores on an abbreviated battery of the Delis-Kaplan Executive Function System. *Archives of Clinical Neuropsychology. 32*(3), 297–305. https://doi.org/10.1093/arclin/acw105

Lovell, M. R., Iverson, G. L., Collins, M. W., McKeag, D., & Maroon, J. C. (1999). Does loss of consciousness predict neuropsychological decrements after concussion? *Clinical Journal of Sport Medicine*, *9*(4), 193-198. https://doi.org/10.1097/00042752-199910000-00002

Lund Research Ltd (2018). *Multiple regression analysis using SPSS statistics.* https://statistics.laerd.com/spss-tutorials/multiple-regression-using-spss-statistics.php

Manchester, D., Priestley, N., & Jackson, H. (2004). The assessment of executive functions: coming out of the office. *Brain Injury*, *18*(11), 1067-1081. https://doi.org/10.1080/02699050410001672387

Marcopulos, B. A., McLain, C. A., & Giuliano, A. J. (1997). Cognitive impairment or inadequate norms: A study of healthy, rural, older adults with limited education. *Clinical Neuropsychologist, 11*(2), 111–131. https://doi.org/10.1080/13854049708407040

McDonald, B. C., Flashman, L. A., & Saykin, A. J. (2002). Executive dysfunction following traumatic brain injury: Neural substrates and treatment strategies. *NeuroRehabilitation*, *17*(4), 333-344. ehttps://doi.org/10m.3233/NRE-2002-17407

Mervis, C.B., Robinson, B.F., & Pani, J.R. (1999). Visuospatial construction. *American Journal of Human Genetics, 65*(5), 1222-1229. https://doi.org/10.1086/302633

Mesulam, M.M. (2000). *Principles of behavioural and cognitive neurology* (2nd ed.). Oxford University Press.

National Health Service (2024). *What are the NHS values?* https://www.nhsprofessionals.nhs.uk/nhs-staffing-pool-hub/working-in-healthcare/what-are-the-nhs-values

Peterson, E., & Welsh, M.C. (2014). The development of hot and cold executive functions in childhood and adolescence: Are we getting warmer? In S. Goldstein & J. Naglieri (Eds.), *Handbook of executive functioning* (pp. 45-65). Springer.

Pickens, S., Ostwald, S. K., Murphy-Pace, K., & Bergstrom, N. (2010). Evidence synthesis: Systematic review of current executive function measures in adults with and without cognitive impairments. *International Journal of Evidence-Based Healthcare*, *8*(3), 110-125. https://doi.org/10.1111/j.1744-1609.2010.00170.x

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills*, *8*(3), 271-276. https://doi.org/10.2466/pms.1958.8.3.271

Ridderinkhof, K. R., Span, M. M., & van der Molen, M. W. (2002). Perseverative behavior and adaptive control in older adults: Performance monitoring, rule induction, and set shifting. *Brain and Cognition*, *49*(3), 382-401. https://doi.org/10.1006/brcg.2001.1506

Roid, G.H. (2003). *Stanford-Binet Intelligence Scales, fifth edition: technical manual.* Riverside Publishing.

Salehinejad, M. A., Ghanavati, E., Rashid, M. H. A., & Nitsche, M. A. (2021). Hot and cold executive functions in the brain: A prefrontal-cingular network. *Brain and Neuroscience Advances*, *5*, 23982128211007769. https://doi.org/10.1177/23982128211007769

Salthouse, T. (1999). Aging and cognition. In R.A., Wilson & F.C. Keil (Eds.*), The MIT encyclopaedia of the cognitive sciences (MITECS)* (pp. 6-7). The MIT Press.

Scheffels, J. F., Ballasch, I., Scheichel, N., Voracek, M., Kalbe, E., & Kessler, J. (2023). The influence of age, gender and education on neuropsychological test scores: updated clinical norms for five widely used cognitive assessments. *Journal of Clinical Medicine*, *12*(16), 1-17. https://doi.org/10.3390/jcm12165170

Schoenberg, M.R., & Scott, J.G. (Eds.). (2011). *The little black book of neuropsychology: A syndrome-based approach.* Springer.

Smith, G.T. (2005). On construct validity: Issues of method and measurement. *Psychological Assessment*, *17*(4), 396-408. https://doi.org/10.1037/1040-3590.17.4.396

Sternberg, R. J., & Powell, J. S. (1983). Comprehending verbal comprehension. *American Psychologist, 38*(8), 878–893. https://doi.org/10.1037/0003-066X.38.8.878

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42(4),* 1096-1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24-31. https://doi.org/10.1177/0098628316677643

Strauss, M.E., & Smith, G.T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology*, *5*(1), 1-25. https://doi.org/10.1146/annurev.clinpsy.032408.153639

Suchy, Y. (2009). Executive functioning: Overview, assessment, and research issues for non-neuropsychologists. *Annals of Behavioral Medicine*, *37*(2), 106-116. https://doi.org/10.1007/s12160-009-9097-4

Tavakol, M., & Wetzel, A. (2020). Factor analysis: A means for theory and instrument development in support of construct validity. *International Journal of Medical Education*, *11*, 245-247. https://doi.org/10.5116/ijme.5f96.0f4a

Tversky, B. (2005). Functional significance of visuospatial representations. In P. Shah & A. Mayake (Eds.), *The Cambridge Handbook of Visuospatial Thinking* (pp. 1-34). Cambridge University Press.

Wechsler, D. (1997). *Wechsler Adult Intelligence eScale-III manual.* The Psychological Corporation

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale--Fourth Edition (WAIS-IV)* [Database record]. APA PsycTests.

Willer, B., & Leddy, J. J. (2006). Management of concussion and post-concussion syndrome. *Current Treatment Options in Neurology*, *8*(5), 415- 426. https://doi.org/10.1007/s11940-006-0031-9

Wilson, A. C. (20234). Cognitive profile in autism and ADHD: A meta-analysis of performance on the WAIS-IV and WISC-V. *Archives of Clinical Neuropsychology.* Advance online publication. https://doi.org/10.1093/arclin/acad073

Wilson, B. A., Alderman, N., Burgess, P. W., Emslie, H., & Evans, J. J. (1996). *BADS: Behavioural Assessment of the Dysexecutive Syndrome*. Pearson.

Zucchella, C., Federico, A., Martini, A., Tinazzi, M., Bartolo, M., & Tamburin, S. (2018). Neuropsychological testing. *Practical Neurology*, *18*(3), 227–237. https://doi.org/10.1136/practneurol-2017-001743

Appendices

Appendix A: Prisma Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	
Study risk of bias	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if	

Section and Topic	Item #	Checklist item	Location where item is reported
assessment		applicable, details of automation tools used in the process.	
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	
Study characteristics	17	Cite each included study and present its characteristics.	
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	

Section and Topic	Item #	Checklist item	Location where item is reported
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	
	23b	Discuss any limitations of the evidence included in the review.	
	23c	Discuss any limitations of the review processes used.	
	23d	Discuss implications of the results for practice, policy, and future research.	
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	
Competing interests	26	Declare any competing interests of review authors.	
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	

Appendix B: Standard and MESH Search Terms

Standard		
Population	Impairment	Intervention Type
"traumatic brain injury" or "acquired brain injury" or "brain injury" or "closed head injury" or "head trauma" or "concussion"	LIST A: "executive function*" or "mental flexibility" or "cognitive flexibility" or "conceptual flexibility" or "metacognition" or "decision making" or "planning" or "problem solving" LIST B: "retrieval" or "recall" or "short term memory" or "working memory" or "metamemory" or or "attention" or "inhibition" or "concept formation" or "abstraction" or "switch*" or "shift*" or "inhibit*" or "initiat*"	"intervention" or "treatment" or "therap*" or "remediation" or "rehabilitation" or "neuropsychological rehabilitation" or "cognitive rehabilitation" or "training" or "neurocognitive training" or "executive function* training" or "neurorehabilitation" or "errorless learning" or "response inhibition"
MeSH		
	LIST A: "executive function" or "metacognition" or "cognitive flexibility" or "decision making" or "problem solving" LIST B: "short term memory" or "human information storage" or "free recall" or "cued recall" or "attention" or "latent inhibition" or "proactive inhibition" or "retroactive inhibition" or "response inhibition or "concept formation" or "abstraction" or "cognitive control" or "reversal shift" or "learning"	"intervention" or "treatment" or "cognitive remediation" or "cognitive rehabilitation" or "neuropsychological rehabilitation" or "training" or "brain training" or "neurorehabilitation" or "learning response inhibition"

Appendix C: Oxford Centre for Evidence-Based Medicine Levels of Evidence

Level	Therapy / Prevention, Aetiology / Harm	Prognosis	Diagnosis	Differential diagnosis / symptom prevalence study	Economic and decision analyses
1a	SR (with homogeneity*) of RCTs	SR (with homogeneity*) of inception cohort studies; CDR" validated in different populations	SR (with homogeneity*) of Level 1 diagnostic studies; CDR" with 1b studies from different clinical centres	SR (with homogeneity*) of prospective cohort studies	SR (with homogeneity*) of Level 1 economic studies
1b	Individual RCT (with narrow Confidence Interval"¡)	Individual inception cohort study with > 80% follow-up; CDR" validated in a single population	Validating** cohort study with good" " " reference standards; or CDR" tested within one clinical centre	Prospective cohort study with good follow-up****	Analysis based on clinically sensible costs or alternatives; systematic review(s) of the evidence; and including multi-way sensitivity analyses
1c	All or none§	All or none case-series	Absolute SpPins and SnNouts" "	All or none case-series	Absolute better-value or worse-value analyses " " " "
2a	SR (with homogeneity*) of cohort studies	SR (with homogeneity*) of either retrospective cohort studies or untreated control groups in RCTs	SR (with homogeneity*) of Level >2 diagnostic studies	SR (with homogeneity*) of 2b and better studies	SR (with homogeneity*) of Level >2 economic studies
2b	Individual cohort study (including low quality RCT; e.g., <80% follow-up)	Retrospective cohort study or follow-up of untreated control patients in an RCT; Derivation of	Exploratory** cohort study with good" " " reference standards; CDR" after derivation, or validated only on	Retrospective cohort study, or poor follow-up	Analysis based on clinically sensible costs or alternatives; limited review(s) of the evidence,

		CDR" or validated on split-sample§§§ only	split-sample§§§ or databases		or single studies; and including multi-way sensitivity analyses
2c	"Outcomes" Research; Ecological studies	"Outcomes" Research		Ecological studies	Audit or outcomes research
3a	SR (with homogeneity*) of case-control studies		SR (with homogeneity*) of 3b and better studies	SR (with homogeneity*) of 3b and better studies	SR (with homogeneity*) of 3b and better studies
3b	Individual Case-Control Study		Non-consecutive study; or without consistently applied reference standards	Non-consecutive cohort study, or very limited population	Analysis based on limited alternatives or costs, poor quality estimates of data, but including sensitivity analyses incorporating clinically sensible variations.
4	Case-series (and poor quality cohort and case-control studies§§)	Case-series (and poor quality prognostic cohort studies***)	Case-control study, poor or non-independent reference standard	Case-series or superseded reference standards	Analysis with no sensitivity analysis
5	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on economic theory or "first principles"

NOTES

Users can add a minus-sign "-" to denote the level of that fails to provide a conclusive answer because:

- ***EITHER*** a single result with a wide Confidence Interval
- ***OR*** a Systematic Review with troublesome heterogeneity.

Such evidence is inconclusive, and therefore can only generate Grade D recommendations.

* By homogeneity we mean a systematic review that is free of worrisome variations (heterogeneity) in the directions and degrees of results between individual studies. Not all systematic reviews with statistically significant heterogeneity need be worrisome, and not all worrisome heterogeneity need be statistically significant. As noted above, studies displaying worrisome heterogeneity should be tagged with a "-" at the end of their designated level.

" Clinical Decision Rule. (These are algorithms or scoring systems that lead to a prognostic estimation or a diagnostic category.)

"¡ See note above for advice on how to understand, rate and use trials or other studies with wide confidence intervals.

§ Met when all patients died before the Rx became available, but some now survive on it; or when some patients died before the Rx became available, but none now die on it.

§§ By poor quality cohort study we mean one that failed to clearly define comparison groups and/or failed to measure exposures and outcomes in the same (preferably blinded), objective way in both exposed and non-exposed individuals and/or failed to identify or appropriately control known confounders and/or failed to carry out a sufficiently long and complete follow-up of patients. By poor quality case-control study we mean one that failed to clearly define comparison groups and/or failed to measure exposures and outcomes in the same (preferably blinded), objective way in both cases and controls and/or failed to identify or appropriately control known confounders.

§§§ Split-sample validation is achieved by collecting all the information in a single tranche, then artificially dividing this into "derivation" and "validation" samples.

" " An "Absolute SpPin" is a diagnostic finding whose Specificity is so high that a Positive result rules-in the diagnosis. An "Absolute SnNout" is a diagnostic finding whose Sensitivity is so high that a Negative result rules-out the diagnosis.

"¡"¡ Good, better, bad and worse refer to the comparisons between treatments in terms of their clinical risks and benefits.

" " " Good reference standards are independent of the test, and applied blindly or objectively to applied to all patients. Poor reference standards are haphazardly applied, but still independent of the test. Use of a non-independent reference standard (where the 'test' is included in the 'reference', or where the 'testing' affects the 'reference') implies a level 4 study.

” ” ” “ Better-value treatments are clearly as good but cheaper, or better at the same or reduced cost. Worse-value treatments are as good and more expensive, or worse and the equally or more expensive.

** Validating studies test the quality of a specific diagnostic test, based on prior evidence. An exploratory study collects information and trawls the data (e.g. using a regression analysis) to find which factors are ‘significant’.

*** By poor quality prognostic cohort study we mean one in which sampling was biased in favour of patients who already had the target outcome, or the measurement of outcomes was accomplished in <80% of study patients, or outcomes were determined in an unblinded, non-objective way, or there was no correction for confounding factors.

**** Good follow-up in a differential diagnosis study is >80%, with adequate time for alternative diagnoses to emerge (for example 1-6 months acute, 1 – 5 years chronic)

GRADES OF RECOMMENDATION

A consistent level 1 studies

B consistent level 2 or 3 studies ***or*** extrapolations from level 1 studies

C level 4 studies ***or*** extrapolations from level 2 or 3 studies

D level 5 evidence ***or*** troublingly inconsistent or inconclusive studies of any level

Appendix D: JBI RCT Checklist

		Choice-Comments/Justification	Yes/No/unclear/NA	Score
Internal Validity				
Bias related to selection and allocation				
1	Was true randomization used for assignment of participants to treatment groups?			
2	Was allocation to treatment groups concealed?			
3	Were treatment groups similar at the baseline?			
Bias related to administration of intervention/exposure				
4	Were participants blind to treatment assignment?			
5	Were those delivering the treatment blind to treatment assignment?			
6	Were treatment groups treated identically other than the intervention of interest?			
Bias related to assessment, detection and measurement of the outcome				
7	Were outcome assessors blind to treatment assignment?			
8	Were outcome measured in the same way for treatment groups?			
9	Were outcomes measured in a reliable way?			
Bias related to participant retention				
10	Was follow up complete and if not, were			

	differences between groups in terms of their follow up adequately described and analysed?			
Statistical Conclusion Validity				
11	Were participants analysed in the group to which they were randomized?			
12	Was appropriate statistical analysis used?			
13	Was the trial design appropriate and any deviations from the standard RCT design (individual randomization, parallel groups) accounted for in the conduct and analysis of the trial?			
Total score				

Appendix E: JBI Quasi-experimental checklist

	Questions	Choices-Comments/Justification	Yes/No/Unclear/NA	Score
1	Is it clear in the study what is the 'cause' and what is the 'effect' (i.e., there is no confusion about which variable comes first)?			
2	Were the participants included in any comparisons similar?			
3	Were the participants included in any comparisons receiving similar treatment/care, other than the exposure or intervention of interest?			
4	Was there a control group?			
5	Was there multiple measurements of the outcome both pre and post the intervention/exposure?			
6	Was follow up complete and if not, were differences between groups in terms of their follow up adequately described and analyzed?			
7	Were the outcomes of participants included in any comparisons measured in the same way?			
8	Were outcomes measured in a reliable way?			
9	Was statistical analysis appropriate?			
Total score				

Appendix F: JBI Case-series Checklist

	Questions	Choices-Comments/Justification	Yes/No/Unclear/NA	Score
1	Were there clear criteria for inclusion in the case series?			
2	Was the condition measured in a standard, reliable way for all participants included in the case series?			
3	Were valid method used for identification of the condition for all participants included in the case series?			
4	Did the case series have consecutive inclusion of participants?			
5	Did the case series have complete inclusion of participants?			
6	Was there clear reporting of the demographics of the participants in the study?			
7	Was there clear reporting of clinical information of the participants?			
8	Were the outcomes or follow up results of cases clearly reported?			
9	Was there clear reporting of the presenting site(s)/clinic(s) demographic information?			
10	Was statistical analysis appropriate?			
Total score				

Appendix G: JBI Checklist Additions

JBI checklist	Question added
JBI checklist for RCTs	Were outcomes measured in a valid way? [test validity] Were outcomes measured in a reliable way? [test reliability]
JBI checklist for quasi-experimental studies	Were the measures used for participants valid?
JBI checklist for case-series	Were measures used for participants valid? Were measures used for participants reliable?

Appendix H: % of studies that found improvement on cold EF outcome measures

Domain	Number of studies addressing this domain	Number of studies that found improvement-time	Percentage of studies that found improvement-time	Number of studies that found improvement-intervention	Percentage of studies that found improvement-intervention
Executive function	4	-	-	2	50%
Problem-solving	5	3	60%	2	40%
Mental flexibility	2	-	-	2	100%
Composite of attention and executive flexibility	2	-	-		100%

Note. Improvement was defined as $p < 0.05$ reported by the authors.

Appendix I: Participant Characteristics

First author (year)	Sample TG/CG	Male/female	CG/TG*	Age range	Time post injury	Severity
Soong (2005)	ABI/ABI	6 male 11 female	5/5/5	Mean = 37.2 SD = - Range = 18-55	Mean: 5.3 years SD: - Qualitative: -	-
Yoshida (2018)	TBI/TBI	16 male 4 female	10/10	Mean = 41.7 SD=9.4 Range = -	Mean: - SD: - Qualitative: At least 6 months	Mild/moderate/ severe
Jacoby (2013)	TBI/TBI	8 male 4 female 1 unknown	6/7	Mean = - SD= - Range = 19-55	Mean: - SD: - Qualitative: At least 2 months	Moderate/severe
Rath (2003)	TBI/TBI	23 male 37 female	28/32	Mean= 43.6 SD = 11.2 Range 22-64	Mean: - SD: - Qualitative: At least 1 year	Mild/moderate/ severe
Fong (2009)	ABI/ABI	27 male 6 female	17/16	Mean = 33.4 SD= 11.5 Range = -	Mean: 12.3 months SD: 13.3 Qualitative: -	Moderate
Man (2013)	TBI/TBI	-	25/25	Mean = - SD= - Range = -	-	Mild/moderate
Sohlberg (2000)	ABI/ABI	-	7 started/7 started	Mean = 35.6 SD= Range = 19-50	Mean: - SD: - Qualitative: At least one year	-
Chen (2011)	ABI/ABI	5 male 7 female	7 started/5 started	Mean: 48 Range: 24-63	Mean: - SD: - Qualitative: At least six months	-
Novakovic-Agopian (2018)	TBI/TBI	30 male 5 female	14/21	Mean =43.3 S=11.57 Range =25-66	Mean: - SD: - Qualitative: At least six months	Mild/moderate/ severe

Marshall (2004)	TBI/TBI	10 male 10 female	20	Mean=35.6 SD=10.9 Range=22-53	Mean: - SD: - Qualitative: At least 20 months	-
Tornås (2006)	ABI/ABI	38 male 32 female	37/33	Mean= 43.0 SD=13.0 Range=19-66	Mean: 97.5 months SD: 112.4 Qualitative: -	-
Thaut (2009)	ABI/ABI	47 male 7 female	23/31	Mean=50.4 SD=- Range=-	Mean: - SD: - Qualitative: At least 14 years	Mild/moderate/ severe

Note. Traumatic Brain Injury; TBI. Acquired Brain Injury; ABI. TG: Target group; CG: Control group; Sample TG/CG = all randomised, not just those who received the intervention. - = not described/unable to determine.

Appendix J: Journal of Clinical and Experimental Neuropsychology Author Guidelines

About the Journal

Journal of Clinical and Experimental Neuropsychology is an international, peer-reviewed journal publishing high-quality, original research. Please see the journal's Aims & Scope for information about its focus and peer-review policy.

Please note that this journal only publishes manuscripts in English.

Journal of Clinical and Experimental Neuropsychology accepts the following types of article: original articles, review articles, critiques.

Preparing Your Paper

Manuscripts submitted to *Journal of Clinical and Experimental Neuropsychology* should conform to the scientific reporting style of the American Psychological Association Publication Manual, 6th Edition (2009).

Structure

Your paper should be compiled in the following order: title page; structured abstract; keywords; main text introduction, materials and methods, results, discussion; acknowledgments; declaration of interest (disclosure) statement; references; appendices (as appropriate); table(s) with caption(s) (on individual pages); figures; figure captions (as a list).

For Results, we ask that you pay special attention to details regarding statistical copy (e.g. rules for numbers expressed as numerals vs. words, decimal fractions, spacing, alignment, and punctuation of statistical copy in text) and formatting of tables. Ensure that tables and figures include a comprehensive key to abbreviations, as well as clear labels for groups, sample sizes, and variables. As per contemporary standards in scientific reporting, effect sizes are an important feature of your report and should be included as appropriate.

Word Limits

Please include a word count for your paper. There are no word limits for papers in this journal.

Style Guidelines

Please refer to these quick style guidelines when preparing your paper, rather than any published articles or a sample copy.

Please use British (-ise) spelling style consistently throughout your manuscript.

Formatting and Templates

Papers may be submitted in Word format. Figures should be saved separately from the text. To assist you in preparing your paper, we provide a formatting template.

Word templates are available for this journal. Please save the template to your hard drive, ready for use.

If you are not able to use the template via the links (or if you have any other template queries) please contact us here.

References

Please use this guide for preparing references in the reporting style of the American Psychological Association.

Checklist: What to Include

1. **Author details.** Please ensure everyone meeting the International Committee of Medical Journal Editors (ICMJE) requirements for authorship is included as an author of your paper. Please ensure all listed authors meet the Taylor & Francis authorship criteria. All authors of a manuscript should include their full name and affiliation on the cover page of the manuscript. Where available, please also include ORCiDs and social media handles (Facebook, Twitter or LinkedIn). One author will need to be identified as the corresponding author, with their email address normally displayed in the article PDF (depending on the journal) and the online article. Authors' affiliations are the affiliations where the research was conducted. If any of the named co-authors moves affiliation during the peer-review process, the new affiliation can be given as a footnote. Please note that no changes to affiliation can be made after your paper is accepted. Read more on authorship.
2. Should contain a structured abstract of **300 words**. Introduction: Describe the background to the study, hypotheses, aims, objectives, research questions, etc. Method: Include outline of the methodology and design of experiments; materials employed and subject/participant numbers with basic relevant demographic information; the nature of the analyses performed. Results: Outline the important and relevant results of the analyses. Conclusions: State the basic conclusions and implications of the study. State, clearly and usefully, if there are implications for management, treatment or service delivery. Note: Any clinical implications should be clearly stated. Avoid abbreviations, diagrams, and references to the text in the abstract.
3. **Graphical abstract** (optional). This is an image to give readers a clear idea of the content of your article. It should be a maximum width of 525 pixels. If your image is narrower than 525 pixels, please place it on a white background 525 pixels wide to ensure the dimensions are maintained. Save the graphical abstract as a .jpg, .png, or .tiff. Please do not embed it in the manuscript file but save it as a separate file, labelled GraphicalAbstract1.
4. You can opt to include a **video abstract** with your article. Find out how these can help your work reach a wider audience, and what to think about when filming.
5. Provide 3 to 6 **keywords**. Ideally 2 keywords will be taken from the NIH MeSH terms (https://meshb.nlm.nih.gov/search). Read making your article more discoverable, including information on choosing a title and search engine optimization. Using accurate MeSH terms as your keywords increases the likelihood that your manuscript, if published, can be identified in searches and increases its impact on the field. Tips: Searching your terms initially allowing for "fragments" versus exact match will generate a list of potential matches to guide your process. NIH U.S. National Library

of Medicine provides "MeSH on Demand", which uses the NLM Medical Text Indexer to identify MESH terms in your submitted text (e.g. abstract or manuscript): https://meshb.nlm.nih.gov/MeSHonDemand. You may also list unique search terms outside the MeSH headings among your keywords.

6. **Funding details.** Please supply all details required by your funding and grant-awarding bodies as follows:
   For single agency grants
   This work was supported by the [Funding Agency] under Grant [number xxxx].
   For multiple agency grants
   This work was supported by the [Funding Agency #1] under Grant [number xxxx]; [Funding Agency #2] under Grant [number xxxx]; and [Funding Agency #3] under Grant [number xxxx].
7. **Disclosure statement.** This is to acknowledge any financial or non-financial interest that has arisen from the direct applications of your research. If there are no relevant competing interests to declare please state this within the article, for example: *The authors report there are no competing interests to declare.* Further guidance on what is a conflict of interest and how to disclose it.
8. **Data availability statement.** If there is a data set associated with the paper, please provide information about where the data supporting the results or analyses presented in the paper can be found. Where applicable, this should include the hyperlink, DOI or other persistent identifier associated with the data set(s). Templates are also available to support authors.
9. **Data deposition.** If you choose to share or make the data underlying the study open, please deposit your data in a recognized data repository prior to or at the time of submission. You will be asked to provide the DOI, pre-reserved DOI, or other persistent identifier for the data set.
10. **Geolocation information.** Submitting a geolocation information section, as a separate paragraph before your acknowledgements, means we can index your paper's study area accurately in JournalMap's geographic literature database and make your article more discoverable to others. More information.
11. **Supplemental online material.** Supplemental material can be a video, dataset, fileset, sound file or anything which supports (and is pertinent to) your paper. We publish supplemental material online via Figshare. Find out more about supplemental material and how to submit it with your article.
12. **Figures.** Figures should be high quality (1200 dpi for line art, 600 dpi for grayscale and 300 dpi for colour, at the correct size). Figures should be supplied in one of our preferred file formats: EPS, PS, JPEG, TIFF, or Microsoft Word (DOC or DOCX) files are acceptable for figures that have been drawn in Word. For information relating to other file types, please consult our Submission of electronic artwork document.
13. **Tables.** Tables should present new information rather than duplicating what is in the text. Readers should be able to interpret the table without reference to the text. Please supply editable files.
14. **Equations.** If you are submitting your manuscript as a Word document, please ensure that equations are editable. More information about mathematical symbols and equations.
15. **Units.** Please use SI units (non-italicized).

Appendix K: Participant Feedback Sheet

Thank you for agreeing to review this study. We are interested to know your opinions on readibility (how easy it is to read), length, format and difficulty.

We will ask you to provide your opinions on:

1. The consent form
2. The participant information form
3. The clinical and demographic questions
4. The psychological test questions

Information sheet	
Was there any language that you didn't understand?	
Were there any words that you think should be changed?	
Was it clear for you what taking part in the study would involve?	
Consent form	
Was there any language that you didn't understand?	
Were there any words that you think should be changed?	
Were there any tick box statements that were not clear?	
Clinical/demographic questions	
Was there any language that you didn't understand?	
Were there any words that you think should be changed?	
Were there any questions that felt intrusive or innapropriate?	

Psychology test	
Were there any test instructions that were not clear/left you feeling confused?	
How did you feel about the length of the test? (e.g. too short, too long, just right)	
Was there any language that you didn't understand?	
Were there any words that you think should be changed?	
These test questions were designed to be 'challenging but passable', e.g. whilst they might make you stop and think, it should be possible to answer them correctly. Did you find the test questions 'challenging but passable'? If not, which questions did you find particularly difficult?	
Do you have any other recommendations?	

Appendix L: Social Media Participant Recruitment Message

Hello all. I am carrying out a research study as part of my DClinPsy and would be incredibly grateful if you would consider taking part (please see the details on the poster below). Anyone between the ages of 11 and 90 who can understand English can take part in this study. It involves answering some online questions. The questions take about twenty minutes. Once you have completed the online questions you will be entered into a prize draw to win a £15 voucher. You have an over 20% chance of winning a voucher. Thanks so much.

Increasing the clinical utility of the SPANS-X: development of the Conceptual Flexibility Index

Background

The SPANS-X is a neuropsychological test. We would like to improve the part of the SPANS-X that provides information about a person's ability to think flexibly or come up with alternative ways of doing things. This improved understanding will help to ensure people get the right kind of clinical help going forward, and the best possible care.

Who can take part?

Anyone between the ages of 11 and 90 who can understand English can take part in this study. You do not need to have a brain injury to take part. You will be asked to complete some questions online, so you will also need to have access to a computer.

What will taking part involve?

You will be asked to complete a series of questions online. These questions should take no more than twenty minutes to complete.

Will I receive anything for taking part?

Once you have completed the online questions you will be entered into a prize draw to win a £15 voucher. You have an over 20% chance of winning a voucher.

Further information and contact details

This study is being organised by Lucy McIvor, a trainee clinical psychologist at CCCU. If you have any questions, would like further information, or would like to take part in the study, please contact Lucy McIvor at **lm1051@canterbury.ac.uk.**

Link to the test: https://www.psytoolkit.org/c/3.4.4/survey?s=W3DeZ

QR CODE:

Appendix M: Participant Demographic and Clinical Questions

What is your first name/nickname?

What is the first letter of your surname?

If you would like to be entered into a prize draw to win a £15 voucher/if you would like to receive a copy of the study results, please enter your email address below so we can contact you. If you do not wish to enter your email address, please click the button to continue.

How old are you?

What was your biological sex at birth?

- ◯ Female
- ◯ Male
- ◯ Intersex

What is your gender?

- ◯ Woman/girl
- ◯ Man/boy
- ◯ Other (please specify)

What is your ethnicity?

- ◯ White British
- ◯ White Irish
- ◯ White Any other White background
- ◯ Mixed White and Black Caribbean
- ◯ Mixed White and Black African
- ◯ Mixed White and Asian
- ◯ Mixed Any other mixed background
- ◯ Asian Indian
- ◯ Asian Pakistani
- ◯ Asian Bangladeshi
- ◯ Asian Any other Asian background
- ◯ Black Caribbean
- ◯ Black African
- ◯ Black Any other Black background
- ◯ Other Chinese
- ◯ Other Any other ethnic group
- ◯ Other Traveller
- ◯ Other Arab

Was 50% or more of your primary education in a Western English speaking country?

- ◯ Yes
- ◯ No

Have you ever in your lifetime got knocked on the head so hard that it knocked you out?

- ◯ Yes
- ◯ No

Has a medical, health or social care professional ever told you that you have an acquired brain injury? (e.g. from an accident or stroke)

- ◯ Yes
- ◯ No

Has a medical, health or social care professional ever told you that you have a cognitive impairment?

- ◯ Yes
- ◯ No

Has a medical, health or social care professional ever told you that you have dementia?

- ◯ Yes
- ◯ No

Has a medical, health or social care professional ever told you that you have a neurodevelopmental condition (e.g. ADHD, autism, dyslexia)?

- ◯ Yes
- ◯ No

Appendix N: University Ethics Approval

This has been removed from the electronic copy.

Appendix O: Feedback to Ethics Panel

Dear Ethics Panel,

Study title: *Increasing the Clinical Utility of the SPANS-X Conceptual Flexibility Index*

I am writing to inform you the above research project has now been completed, and a thesis has been written to be submitted for partial fulfilment of the degree of Doctor of Clinical Psychology at Canterbury Christ Church University.

Impairments in 'cold' executive functioning abilities following acquired brain injury are frustrating and inconvenient and can negatively impact on quality of life. At present, a wide range of neuropsychological tests exist to assess for impairments in cold executive functioning. From a clinical perspective, this can mean that clients are required to complete a vast battery of neuropsychological tests to assess for possible impairments, which can be uncomfortable and anxiety provoking. It also increases the likelihood that clients will obtain at least one 'impaired' score simply by chance.

One solution to these dilemmas is to 'narrow' the construct of cold executive functioning, through focusing on a specific facet of the construct. The Short Parallel Assessments of Neuropsychological Status-X (SPANS-X) achieves this narrower focus via a conceptual flexibility index (CFI). However, the CFI of the SPANS-X required development to remain clinically useful. This study aimed to develop the CFI of the SPANS-X via improving its construct/convergent validity and internal consistency reliability. In order to achieve this, new theoretically-driven items and subtests were added to the existing SPANS-X CFI. This expanded SPANS-X CFI was then digitalized and completed by participants.

The sample comprised of 50 participants aged 20-87 years (76% female, 52% university educated, 84% White British, 76% reporting no concussion, 80% reporting no neurodiversity). A non-experimental, cross sectional survey design was used. Participant scores on the newly added items and subtests were correlated with their scores on the original CFI to assess convergent validity, whilst Cronbach's alpha coefficients were used to evaluate internal consistency reliability. A multiple linear regression was also conducted to explore the impact of sex, age, education level, concussion, and neurodiversity on CFI score.

The results showed that certain newly added items and subtests showed 'good' convergent validity with the original CFI (0.51). Certain newly added items and subtests also increased the internal consistency reliability of the CFI from 'acceptable' to 'good' (0.72 to 0.83). These positive findings tentatively lend credence to the existence of 'conceptual flexibility' as a sub-construct of cold executive functioning and support the inclusion of these items and subtests within further test validation studies.

In contrast, other items and subtests did not appear to increase the validity or reliability of the existing SPANS-X CFI. This may be because they demonstrated too much response variation, or 'loaded' onto theoretical constructs other than conceptual flexibility, such as visuospatial functioning or verbal reasoning. Another possibility is that items and subtests were not pitched at the appropriate difficulty level or were not appropriately worded or scored. These items and subtests should not therefore advance to further testing phases.

In terms of the influence of contextual variables, age and education level had a significant impact on CFI score, suggesting that these characteristics are important to account for when interpreting client performance. Mean CFI score tended to decrease with age. This accords with research suggesting that, in line with the natural ageing process, mean neuropsychological test

scores tend to decrease with age amongst healthy adults. Mean CFI score tended to increase with education level. This accords with research suggesting that those who are well educated may have had enhanced opportunities to acquire skills that make them more successful when completing psychometric tests.

The aim of this research was to contribute to the development of the SPANS-X CFI, enhancing its clinical utility. On a wider scale, it is hoped that this positively contributes to the development of neuropsychological tests which specifically focus on a sub-construct of cold executive functioning. This in turn could improve clients' experience of neuropsychological testing by reducing the number of tests they are required to undertake in clinical settings, alleviating discomfort and anxiety and decreasing the likelihood that clients will score an 'impaired' score by chance. In terms of future research, it will be important to repeat the neuropsychological test designed for this study with a wider sample of participants, extending beyond healthy populations to participants who have sustained an acquired brain injury or are diagnosed with other neurological conditions.

Appendix P: Feedback to Participants

Thank you again for taking part in research about the SPANS-X.

The SPANS-X is a neuropsychological test. The SPANS-X can provide clinicians with a range of information about the way in which a person's brain is working after an acquired brain injury. The aim of the research was to improve the part of the SPANS-X that provides information about a person's ability to think flexibly and come up with alternative ways of doing things. We aimed to improve this part of the SPANS-X by adding extra questions and subtests. You completed the 'expanded' version of the SPANS-X online.

We have now analysed the results. The results are based on 50 participants aged 20-87 years (76% female, 52% university educated, 84% White British, 76% reporting no concussion, 80% reporting no neurodiversity**).**

Summary of the results:

We found that certain questions appeared to improve the part of the SPANS-X that provides information about a person's ability to think flexibly. These questions included:

- *Parts of the 3-and-1 concept test (where you were asked to identify similarities and differences between visual objects)*
- *Parts of the similarities test (where you were asked to identify what two objects had in common)*
- *Parts of the association test (where you were asked to explain the verbal relationship between objects)*
- *Parts of the proverbs A test (where you were asked to explain the meaning of a proverb)*

We found that other questions did not appear to improve the part of the SPANS-X that provides information about a person's ability to think flexibly. These questions included:

- *Parts of the 3-and-1 concept test*
- *Parts of the similarities test*
- *Parts of the association test*
- *Parts of the proverbs A test*
- *All of the proverbs B test (where you were asked to select the meaning of a proverb from a multiple-choice list)*

We also found that a person's age and a person's level of education tended to influence their score on the expanded version of the SPANS-X. On average, as age increased, score decreased. On average, as level of education increased, score increased.

What are the benefits of the research?

- The results show that adding certain tests and questions may improve the part of the SPANS-X that provides information about a person's ability to think flexibly
- This in turn may help us gain a better understanding of a person's ability to think flexibly after an acquired brain injury
- This enhanced understanding will help to ensure that people who have an acquired brain injury get the right kind of clinical help going forward, and the best possible care

Appendix Q: Participant Information Sheet

Participant Information Sheet

Increasing the clinical utility of the SPANS-X: development of the Conceptual Flexibility Index

Summary

Hello, my name is Lucy McIvor and I am a trainee clinical psychologist at Salomons Institute for Applied Psychology, Canterbury Christ Church University. This research project is being carried out as part of my professional qualification. My supervisor is Dr Jerry Burgess (jerry.burgess@canterbury.ac.uk). Dr Burgess is a clinical neuropsychologist. We would like to invite you to take part in our study. The study aims to improve the SPANS-X. The SPANS-X is a neuropsychological test.

Background to the research

The SPANS-X is a neuropsychological test. The SPANS-X can provide clinicians with a range of information about the way in which a person's brain is working. We would like to improve the part of the SPANS-X that provides information about a person's ability to think flexibly or come up with alternative ways of doing things. A better understanding of a person's ability to think flexibly or come up with alternative ways of doing things will help to ensure they get the right kind of clinical help for their brain injury going forward, and the best possible care.

Who can take part in this study?

Anyone between the ages of 11 and 90 who understands English can take part in this study. You do not need to have a brain injury to take part. You will be asked to complete some questions online, so you will also need to have access to a computer.

What would taking part involve?

If you would like to take part in the study, you will be asked to fill out an online consent form. This can be signed digitally. You will also be asked questions about your age, sex at birth, gender, ethnicity, home residence, language status, schooling information and education level. We will also ask you if you have a brain injury or a neurodevelopmental condition. This is to help ensure the results from our study are relevant to a wide range of people. You will then be asked to complete a series of online questions. These questions should take no more than 20 minutes to complete.

What are the possible benefits of taking part?

We hope that if we are able to improve the SPANS-X, we can improve the care provided to people with brain injuries.

What are the possible disadvantages and risks of taking part?

If you have a brain injury or a neurological condition, you might find it hard to complete a test that reminds you of this is in a negative way. There is information at the end of this sheet should you wish to find further help and support.

What will happen if I don't want to carry on taking part?

You can stop being part of the study without giving a reason up to 72 hours after you have completed the study by informing Lucy McIvor (lm1051@canterbury.ac.uk). We will assume that we can keep the anonymous information and scores we already have about you.

How will you use information about me?

We will need to use information from you for this research project. This information will include your name, age, sex at birth, gender, ethnicity, home residence, language status, schooling information and education level. We will also need to know if you have a brain injury or a neurodevelopmental condition. We will also ask you to provide your email address if you would like the chance to win a voucher. People who do not need to know who you are will not be able to see your name or email address. Your information will be made anonymous at the first opportunity through assigning you a coded number to be used instead of your name. We will keep all information about you safe and secure on a password protected spreadsheet. Once we have finished the study, your information will be stored in a password protected file in a locked office for 10 years and then then destroyed. We will write our study in a way that means no-one can work out that you took part in the study.

What will happen to the results of this study?

We hope to circulate results internally within services and at scientific conferences. We also hope to publish the results of this study in peer-reviewed scientific journals.

Can I access the results of the study?

If you would like to see the final results of the study, we would be pleased to send the completed research to your email if you provide it. Unfortunately, we cannot provide results on individual performance (e.g. what you scored). This is because we don't yet know the accuracy of the test questions you will complete, so we cannot provide you with any clinical information or any possible diagnosis. If you are worried about your health, you should speak to a clinical professional outside of research.

Who is organising and funding the study?

This study is being organised by Lucy McIvor, a trainee clinical psychologist at Canterbury Christ Church University and is supervised by Dr Jerry Burgess, a clinical neuropsychologist at Canterbury Christ Church University. The data protection officer is Dr Fergal Jones, Canterbury Christ Church University.

Will I receive anything for taking part in the study?

Once you have completed the online questions you will be entered into a prize draw to win a £15 voucher. You have an over 20% chance of winning a voucher.

Who is the Chief Investigator of the study?

The chief investigator for the study is Lucy McIvor (lm1051@canterbury.ac.uk)

Who has reviewed this study?

This study has been approved by the Salomons Research Ethics committee at Canterbury Christ Church University.

Further information and contact details

If you have any questions, would like further information, or would like to take part in the study, please contact Lucy McIvor at lm1051@canterbury.ac.uk.

Where can I access further support if needed?

You may wish to contact your GP in the event of urgent need.

Live well: A comprehensive NHS website that provides advice for general and specific mental health issues and advice on self-care.

https://www.nhs.uk/live-well/

Mind: A website useful for those experiencing mental distress.

https://www.mind.org.uk/

Mental Health Foundation: Free mental health podcasts aimed at improving mental health and tackling anxiety, fear, stress, anger, physical health issues and poor sleep.

https://www.mentalhealth.org.uk/podcasts-and-videos/podcast-interviews

Samaritans: free confidential support 24 hours a day.

https://www.samaritans.org/how-we-can-help/contact-samaritan/

Appendix R: Participant Consent Form

Confirm you want to do this survey

Please read and tick the statements below to confirm you are happy to participate in this study.

- [ ] 1. I confirm that I have read the information sheet for the study. I have had the opportunity to consider the information, ask questions and gain satisfactory answers.
- [ ] 2. I understand that my participation is voluntary and that I am free to withdraw without giving a reason up to 72 hours after my participation in the research is complete.
- [ ] 3. I understand that if I withdraw, any information I have already provided will be kept in anonymised form.
- [ ] 4. I understand that I will not be entered into the gift voucher draw unless I complete the series of online questions.
- [ ] 5. I understand that my identity will remain anonymous in the analysis and publication of the research.
- [ ] 6. I understand any sensitive and identifiable information I provide will be held in strict confidentiality by Salomons Institute of Applied Psychology, Canterbury Christ Church University. Identifiable information e.g. name will be retained separately from test result data and other information.
- [ ] 7. I understand that after completion of the study, my anonymised data will be kept in a password protected file to be stored in the Institute office in a locked cabinet for ten years and then destroyed.
- [ ] 8. I agree to take part in the above study.

Appendix S: Power Analysis Calculations

Table S1

Power Calculation for Spearman Correlation Tests

				Test assumptions		
	N	Actual power[b]	Power	Null	Alternative	Significance
Spearman Correlation[a]	27	0.81	0.80	0	0.50	0.05

a. One-sided test.
b. Based on Fisher's z-transformation and normal approximation. The variance estimation is based on the method suggested by Bonett and Wright.

Table S2

Power Calculation for Multiple Linear Regression Analysis

			Predictors		Test assumptions		
	N	Actual power[b]	Total	Test	Power	Partial[c]	Significance
Type III F-test[a]	45	0.81	5	5	0.80	0.50	0.05

a. Intercept term is included.
b. Predictors are assumed to be fixed.
C. Multiple partial correlation coefficient.

Appendix T: Statistical Assumptions of Multiple Linear Regression Analysis

The assumptions of a multiple linear regression are as follows (Lund Research Ltd, 2018):

The dependent variable should be measured on a continuous scale	The dependent variable, expanded SPANS-X CFI score, was measured on a continuous scale.
Independent variables should be measured on either a continuous or categorical scale, and show a linear relationship to the dependent variable	All independent variables were either continuous or categorical. The variables age and education level demonstrated a linear relationship with the independent variable as evidenced by scatter plots **(Figures T1 & T2).** It was not necessary to check the linearity assumption for the categorical predictors, as the linearity assumption is always met as a straight line always fits two points (each category) exactly.
The data should show independence of residuals, as evidenced by a Durbin-Watson statistic between 1.5 and 2.5.	The data showed independence of residuals, as evidenced by a Durbin-Watson statistic of 1.95.
Variables should not violate the assumption of homoscedasticity, which is assessed by plotting the regression standardized residuals (Y axis) against the regression predicted values (X axis). If data shows homoscedacity, there should be no pattern to the scatter as evidenced by a best fit line	The variables did not violate the assumption of homoscedacity, as demonstrated by a straight best fit line **(Figure T3).**
Data must not show multicollinearity, which occurs when two or more of the independent variables inputted into the regression are highly correlated. Multicollinearity may lead to problems understanding which independent variable contributes to variance observed in the dependent variable. A variance inflation factor (VIF) < 5 indicate there is no multicollinearity.	The data did not demonstrate multicollinearity, as no VIF values > 5 **(Table T1).**
There should be no significant outliers or highly influential data points, as these points may have a disproportionate effect on the regression equation. A Cook's distance greater than 1 indicates a significantly influential data point.	No Cook's distance values were greater than 1, indicating there were no significantly influential data points
Regression residuals (errors) should be approximately normally distributed. This may be evidenced through the use of histograms and P-P plots.	The histogram and P-P plots show that residuals are approximately normally distributed **(Figures T4 & T5).**

Figure T1

A Scatterplot to Demonstrate the Relationship between Expanded CFI Score and Age

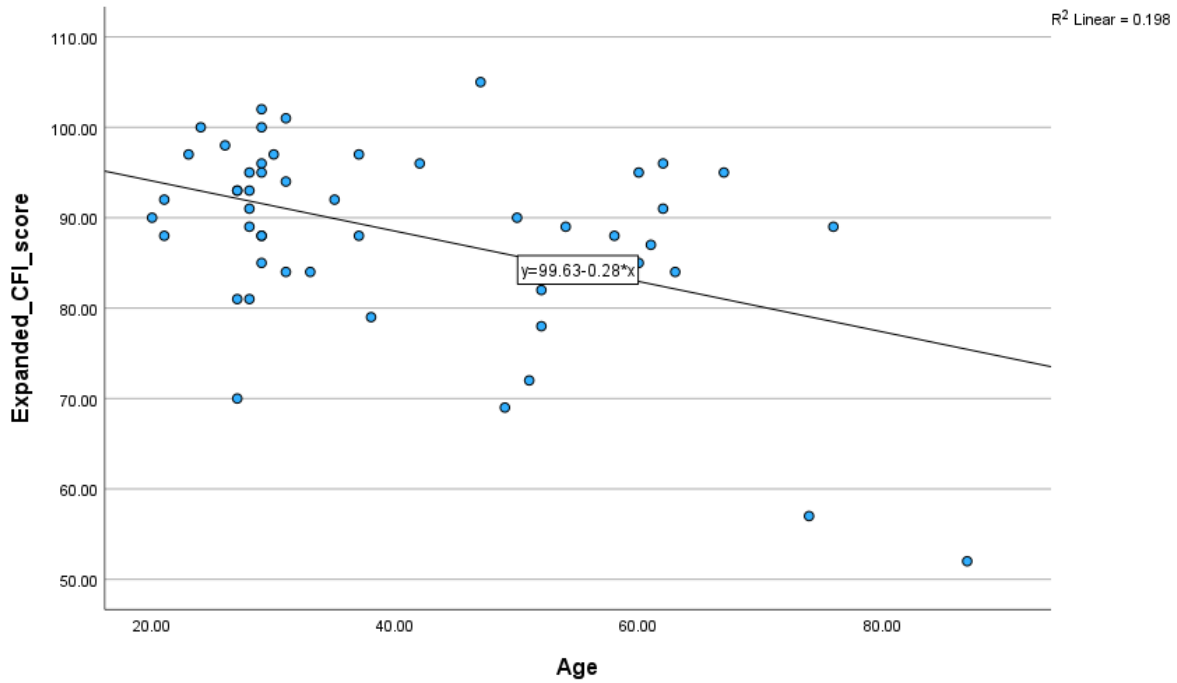


Figure T2

A Scatterplot to Demonstrate the Relationship between Expanded CFI Score and Education

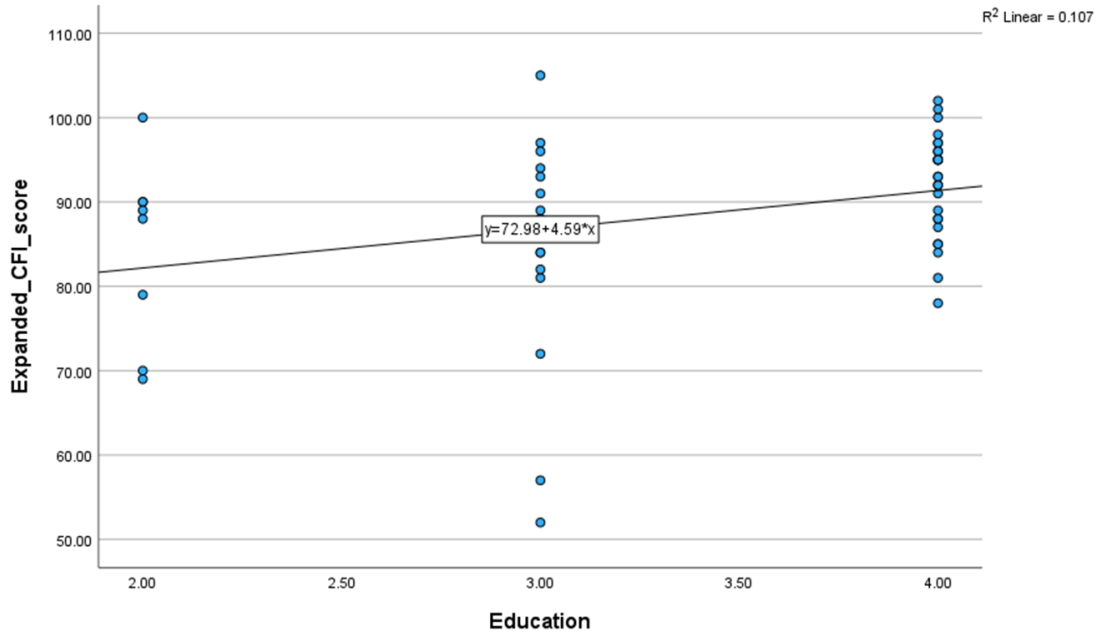


Figure T3

A Scatterplot to Demonstrate the Homoscedastic Relationship between the Regression Predicted Values and Regression Standardized Residuals

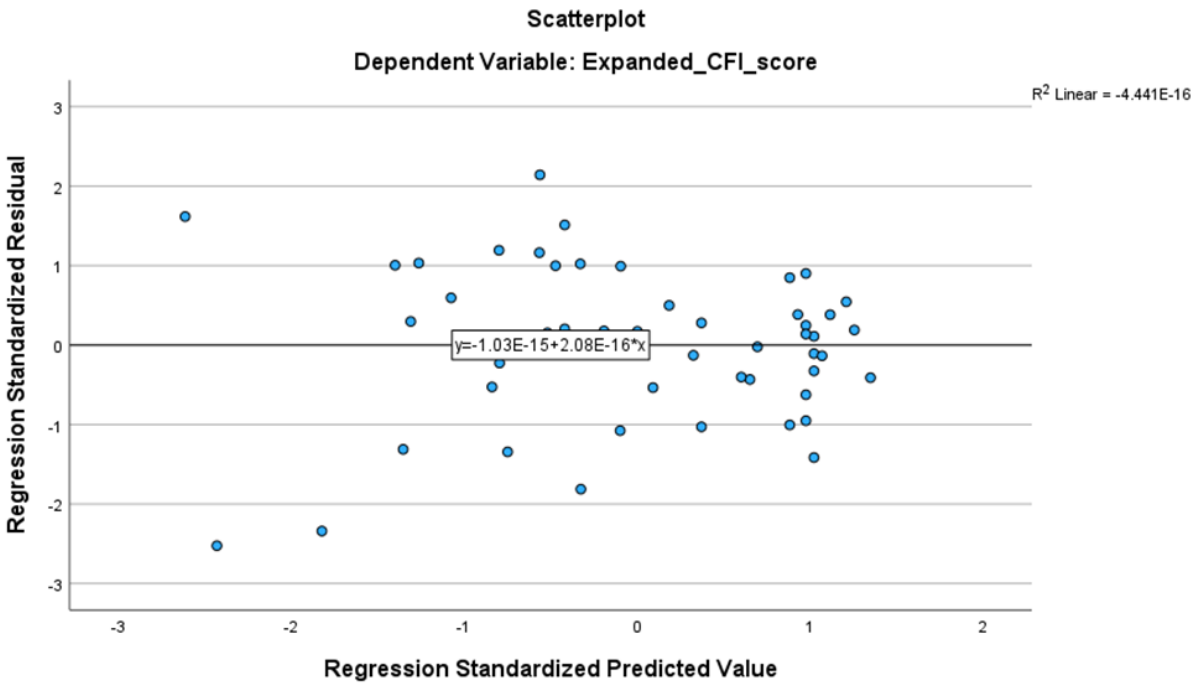


Table T1

Variance Inflation Factor Values for Each Predictor Variable within the Regression Model

Variable	VIF
Age	1.02
Education	1.02
Sex	1.01
Concussion?	1.05
Neurodiversity?	1.03

Figure T4

A Histogram to show the Distribution of Regression Residuals

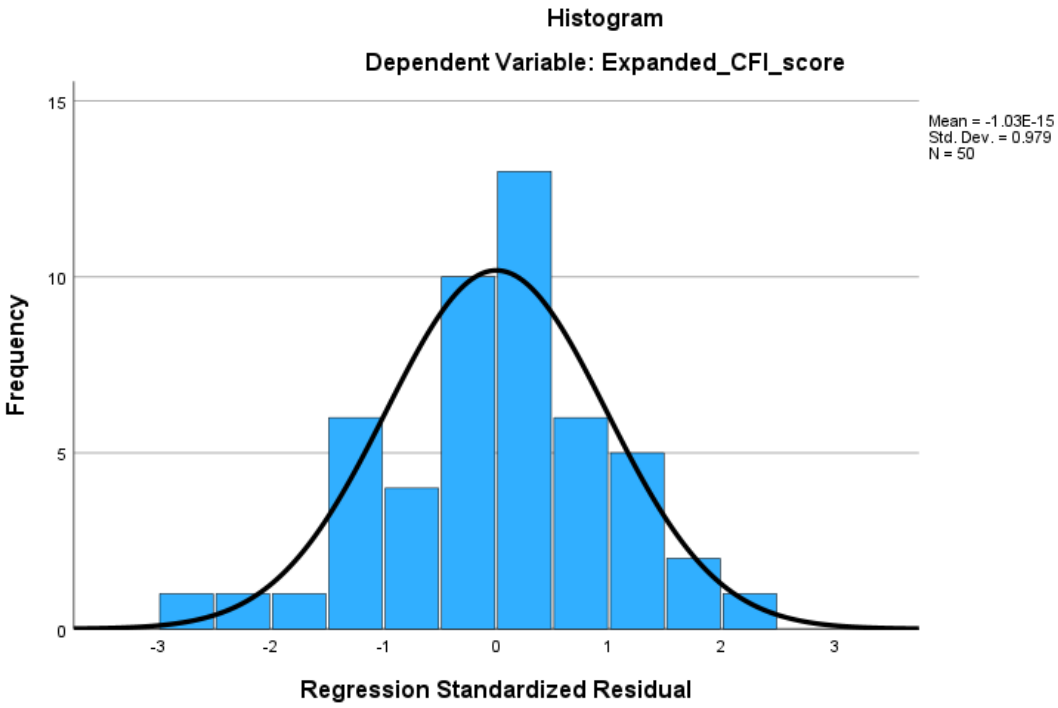


Figure T5

A P-P Plot to show the Distribution of Regression Residuals